Data for Accountability in Education

NORRAG SERIES ON INTERNATIONAL EDUCATION AND DEVELOPMENT

Series Editors: Moira Faul and Chanwoong Baek, *Network for International Policies and Cooperation in Education and Training (NORRAG)*

The NORRAG Series on International Education and Development reflects the mission of NORRAG, the Global Education Centre of the Geneva Graduate Institute and Global Network for Policies and International Cooperation in Education and Training, to surface underrepresented expertise to support diverse stakeholders in addressing the global complexities of education and development. We produce, disseminate and broker critical knowledge on topical issues that emerge in education and development. Through its programs, knowledge production and dissemination, NORRAG contributes to enhancing the conditions for participatory, informed and evidence-based policy decisions that improve equity and quality of education. This series aims to broker knowledge at the interface between research, analysis, policy and practice within the comparative, development and international education community.

For a full list of Edward Elgar published titles, including the titles in this series, visit our website at www.e-elgar.com.

Data for Accountability in Education

Global Trends in School Reform

Gita Steiner-Khamsi

William H. Kilpatrick Professor, Teachers College, Columbia University, USA

Patricia Bromley

Associate Professor, Stanford University, USA

Rie Kijima

Assistant Professor, University of Toronto, Canada

Kerstin Martens

Professor, University of Bremen, Germany

Antoni Verger

Professor, Autonomous University of Barcelona, Spain

NORRAG SERIES ON INTERNATIONAL EDUCATION AND DEVELOPMENT

Edward Elgar PUBLISHING

Cheltenham, UK · Northampton, MA, USA

Cover image: Designed by Anouk Pasquier (source file from starline / Freepik)

Published by

Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

Authorised representative in the EU for GPSR queries only: Easy Access System Europe – Mustamäe tee 50, 10621 Tallinn, Estonia, gpsr.requests@easproject.com

A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2025951271

This book is available electronically in the Elgaronline Sociology, Social Policy and Education subject collection
https://doi.org/10.4337/9781035349869

ISBN 978 1 0353 4985 2 (cased)
ISBN 978 1 0353 4986 9 (eBook)
ISBN 978 1 0353 9497 5 (ePub)

Contents

List of figures

List of boxes

About the authors

Patricia Bromley is Associate Professor of Education, Environmental Social Science and (by courtesy) Sociology at Stanford University. Her research spans a range of fields, including comparative education, organisation theory, sociology of education, and public administration and policy. Her work focuses on the historical rise and globalisation of liberal culture emphasising rational, scientific thinking and expansive forms of rights as well as contemporary attacks on this culture. Much of her research empirically focuses on two settings – education systems and organisations. Her recent publications examine pushbacks against liberal culture in the form of growing restrictions on civil society and declining emphases on education reform.

Rie Kijima is Assistant Professor at the Munk School of Global Affairs and Public Policy, University of Toronto, where she serves as the inaugural director of the Initiative for Education Policy and Innovation. Rie Kijima's research addresses topics such as the politics of international assessments, the impact of education reforms, and STEAM education. Her recent articles have appeared in journals such as *The Review of International Organization*, *Review of International Political Economy*, *Sociology of Education* and *International Journal of STEM Education*. Her research project on education reform is supported by the Social Sciences and Humanities Research Council of Canada. She received her PhD and MA from Stanford University and her BA from the ICU in Tokyo, Japan.

Kerstin Martens is Professor of International Relations at the University of Bremen, Germany. She is head of the research project A05, "The Global Development of Coverage and Generosity in Public Education", at the Collaborative Research Center 1342 "Global Dynamics of Social Policy", also based at the University of Bremen. She has published widely on international organisations in education policy (particularly on

the OECD and the UN), international large-scale assessments in education, such as the PISA study, and on global public policies, particularly those focusing on education, health and social policy. She holds a PhD from the European University Institute in Florence, Italy.

Gita Steiner-Khamsi is the William H. Kilpatrick Professor of Comparative Education at Teachers College, Columbia University, New York. She also holds the Honorary UNESCO Chair in Comparative Education Policy at the Geneva Graduate Institute. A former president of the Comparative and International Education Society (2009/2010), editor of the *World Yearbook of Education* (2008–2021) and academic director of NORRAG (2017–2022), she has published widely on topics such as policy transfer, qualitative comparative analyses and school reform from an international comparative perspective. Her most recent book, *Time in Education Policy Transfer: The Seven Temporalities in Global School Reform*, was published with open access in English (Palgrave/Springer Nature), Chinese (Hong Kong University Press) and Spanish (Prometeo).

Antoni Verger is Professor of Sociology at the Autonomous University of Barcelona and an ICREA research fellow. His research focuses on the relationship between governance institutions and education policy, using comparative policy analysis to examine the adoption, trajectory and impact of key education reforms, including public–private partnerships, accountability mechanisms and school governance models. He has secured multiple research grants on these topics, including two from the European Research Council for the REFORMED and ACCOUNTED projects. He also holds editorial leadership roles with the *World Yearbook of Education* and the *Journal of Education Policy*. In addition, he directs the Erasmus+ joint master's programme Education Policies for Global Development (GLOBED), which has received three awards from the European Commission for its excellence in international education.

Preface

Over the past few years, each of us has researched the book's topic within our own networks. We have taught courses on datafication and accountability trends, enlisted our dedicated students, recruited like-minded colleagues, and presented our research at academic conferences and in journals of our choice. Occasionally, our paths crossed, creating opportunities for lively debates that we valued. These intellectually stimulating encounters prompted us to consider whether we should collaborate and share our insights under the same book cover.

A cooperative effort like this is uncommon due to our varying interpretive frameworks (institutionalist theory, political economy, sociological systems theory), methodological approaches along the qualitative–quantitative spectrum, and diverse disciplinary backgrounds (organisational sociology, political economy, international relations, public policy studies). This book emerged from multiple meetings where we learned from each other, discussing how our approaches complement one another to provide a more comprehensive understanding of how data-for-accountability has evolved over time, the factors and drivers behind these changes, and the impacts these changes have had on students, educators, schools, governments, and global actors.

The book benefited from several ongoing research initiatives, including the REFORMED project (grant number 680172), funded by the European Research Council under the European Union's "Horizon 2020 Framework Programme for Research and Innovation" and based at the Autonomous University of Barcelona. Several authors of this book were also able to draw on the rich information compiled in the World Education Reform Database (WERD). The WERD research initiative is funded by the Social Sciences and Humanities Research Council of Canada and Stanford's Center on Philanthropy and Civil Society. It is based at the Stanford Graduate School of Education and the Munk School of Public Policy at the University of Toronto. Finally, the book gained from the studies carried out in the Collaborative Research Center 1342 "Global Dynamics of Social Policy" at the University of Bremen.

The centre is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), and the assigned project number is 374666841—SFB 1342.

A note of gratitude is in order here. Courtney Martinez, Doctoral Fellow at Teachers College, Columbia University, provided invaluable support for formatting the manuscript. We are grateful to the Geneva Graduate Institute, in particular the NORRAG Global Education Centre for including our book in their open-access book series. Chanwoong Baek, Academic Director of NORRAG and a faculty member in the Department of Political Science and International Relations at the Geneva Graduate Institute, has accompanied and enthusiastically supported the book project from its inception to its completion.

Gita Steiner-Khamsi, Teachers College, Columbia University, New York
Patricia Bromley, Graduate School of Education and Doerr School of Sustainability, Stanford University
Rie Kijima, Munk School of Global Affairs and Public Policy, University of Toronto
Kerstin Martens, Institute for Intercultural and International Studies, University of Bremen
Antoni Verger, Department of Sociology, Autonomous University of Barcelona

1. Understanding global reform trends and their varied impacts: making a case for complementary research perspectives and agendas

Gita Steiner-Khamsi, Patricia Bromley, Rie Kijima, Kerstin Martens and Antoni Verger

Data *on, for* and *in* education has rapidly accumulated since the turn of the millennium. Over 2,500 internationally comparable indicators of education currently exist (World Bank, 2024).[1] In 2024, the UNESCO Institute for Statistics expanded the Sustainable Development Goal 4 indicators and added 30,041 new data points on other policy-relevant indicators (UIS, 2024). Intergovernmental organisations are not the only organisations prioritising data gathering in their daily operations. Data collection by groups advocating for more, better or more equitable education has also surged. International advocacy groups such as NORRAG and the Inter-Agency Network for Education in Emergencies use various types of quantitative data to argue for counting the uncounted (e.g., internally displaced persons) when improving access to education. Finally, data mining in education involves extracting information on students, schools, districts, provinces, countries, regions and the world. The growing demand for data has transformed education into a measurable, calculable and seemingly predictable unit of analysis.

This book examines the global trend of utilising data for accountability purposes. Accountability is an inherently relational concept, as

[1] These indicators cover areas such as access, progression, completion, literacy, teachers, population and expenditure.

it requires one entity to report to another and, conversely, one entity to monitor the other. Examples of accountability include international organisations monitoring countries' educational progress, schools reporting to educational authorities and NGOs overseeing school providers or the use of public education funds, among others. The dynamics surrounding who reports to whom and who monitors whom have not only evolved over time but also differ across regions and countries.

The term "data" in this work refers to formalised administrative processes and systematically collected observations for the purpose of analyzing or describing aspects of schooling to render education visible. Data in quantitative formats, such as enrolment figures or test scores, have taken on a monumental role. In today's world, recent advances in computing and artificial intelligence have added a new layer of complexity to quantitative forms of data (e.g. predictive behaviour). Data can also include qualitative descriptions, for example, in formal reports submitted to national governments or international organisations (e.g. narrative judgements about leadership or teaching, and education sector analysis reports). In contrast, a school inspector visiting a rural school and verbally recommending that a central authority keep the school open would not be considered data because it leaves no formal record or traceable evidence.

In this book, we explore and clarify the processes and outcomes of the global trend of using data for accountability purposes. We trace this trend back to the 1990s, a key inflection point in its development, and demonstrate how an ever-growing number of actors have employed it, tailored it to their purposes and, as a result, diversified its utility as a governance tool.

1. THE DATA-FOR-ACCOUNTABILITY REFORM: CONSIDERATIONS OF PROCESS AND OUTCOMES

The data-for-accountability approach has proven versatile, with each new country selectively adopting elements of the "governance-by-numbers" toolkit. Numerous tool variants emerged, among which large-scale assessments stand out, lowering the threshold for subsequent adopters to engage in the global movement. Thus, inflationary use has been both a consequence of and a catalyst for the rapid pace of global diffusion. Intergovernmental organisations actively supported this trend by offering advice and, in some cases, funding for the adoption of new assessment

frameworks and data infrastructures. Such organisations also facilitated the use of data for reporting and monitoring purposes by proposing international declarations and agreements with measurable targets and benchmarks, such as the World Declaration on Education for All and, later, the Millennium Development Goals. A busy influx of global reforms has been introduced worldwide due to the continuously increasing number of international agreements, many of which include development bencharmarks and targets that signatories are expected to meet. The number of international agreements registered with the United Nations grew from 8,776 in 1960 to 63,419 by March 2010, reflecting how international institutions have progressively expanded their mandates and legal instruments across an increasing number of governance areas, such as human rights, environmental issues and trade treaties (Zürn, 2018).

When, why and how policymakers selectively adopt data-for-accountability as a governance tool are equally essential lines of enquiry. Several political constellations and institutional legacies remind us that this reform approach was not inscribed onto an empty canvas. Policymakers often reframed the tool as a (global) solution to a (local) problem. In the wake of this creative act, policymakers have had to distance themselves from past solutions and current alternatives to gain leverage for their reform endeavours. Any investigation into the processes of global diffusion, national trajectories and institutional path dependencies arguably needs to pay close attention to feedback loops. Although understudied, feedback loops exist between early and late adopters (governments, educational authorities), between adopters and implementers (classrooms, schools and districts), and between adopters and global institutions (the OECD, the World Bank and many other international and regional organisations). An in-depth scrutiny of existing and missing feedback loops helps us understand why some governance tools, such as data-for-accountability purposes, have become both more ubiquitous and more elusive over time. Several scholars observed, for example, that the World Bank does not consider feedback from local actors (Diaz Rios & Urbano–Canal, 2023) and continues to disseminate its portfolio of best practices globally in a self-referential manner (Seitzer, Baek & Steiner-Khamsi, 2025). Similarly, late adopters of the government downsizing reform of the 1980s and 1990s were not keen to learn about the negative impact of the reform on early adopters (Lee & Strang, 2006). So-called vicarious learning only applies to early adopters: the early adopters of a reform are eager to learn from the experiences of elsewhere, whereas late

adopters join a global reform movement for symbolic reasons or due to normative emulation pressures (Dobbin et al., 2007).

In addition to investigating processes, we offer several interpretive approaches to seeing and understanding the policy outcomes of data-driven governance tools. As with our multilevel analyses of processes, we observe varied outcomes depending on who uses the tool for which purposes: global actors, national policymakers, school administrators, parents or students. To illustrate this point, we may compare two groups that are often portrayed as occupying opposite ends of the governance spectrum: those who govern (global actors and educational authorities) and those who are governed (school actors, including teachers, principals and students) - although in practice, relationships within and across these levels are characterised by mutual interdependence and shifting internal hierarchies.

International organisations, on what seems like a core and routinary process, develop new indicators that are made for both analysis and broad communication - e.g. in the form of indexes, scores, ranks, ratios or percentages. For example, the "learning loss" is one such indicator that was widely discussed and spread towards the end of the COVID-19 pandemic to capture the impact of school closures, remote learning, and alternative forms of schooling on learning. Likewise, the concept of "learning poverty" has gained prominence in recent years. Like the Human Capital Index, the learning poverty index amalgamates attainment (enrolment) and achievement (test scores) into a single measure. Because these indices bundle access, learning quality, and health (in the case of HCI), their scores often diverge from simple measures of school enrolment. Far from neutral, these composite indicators carry normative and methodological assumptions — about what counts as "learning," which skills matter, and how to compare across diverse contexts — and thus operate not only as descriptive tools but as instruments of norm-setting and policy framing. The deficit-oriented learning poverty index, for example, contributes to the construction of a worldwide "learning crisis," as reflected in the following quotation by one of its main proponents:

> We have determined that 53 percent of children in low- and middle-income countries cannot read and understand a simple story by the end of primary school. In poor countries, the number is as high as 80 percent. Such high levels of illiteracy are an early warning sign that all global educational goals and related sustainable development goals are in jeopardy. (World Bank, 2021)

This framing places the burden of accountability on governments and urges them to take responsibility for poor results while simultaneously encouraging them to intensify accountability mechanisms within their own education systems. This turn towards quantifiable measures that can be used for public accountability purposes is part of a wider datafication of education (and broader society) – from a proliferation of standardised student tests to a flurry of monitoring, evaluation and reporting requirements throughout all levels of the education system. As Nóvoa and Yariv-Mashal (2003) noticed in the early intensification of contemporary datafication, more than 20 years ago, "[T]his ongoing collection, production, and publication of surveys leads to an 'instant democracy', a regime of urgency that provokes a permanent need for self-justification" (p. 427).

Today, three decades after numerous public administrations worldwide committed to making greater use of research evidence, we possess more data than ever; despite this, there is a lack of uptake in policy, planning and practice. The so-called "surplus of evidence" (Lubienski, 2019) is a direct outcome – or a logical consequence – of the quality assurance measures implemented in the last decades. Nonetheless, in some countries it has evolved alongside–and partly through–the outsourcing of school provision to non-state actors. Henceforth, data were used to hold both private and public actors accountable for the quality of their performance. Within a short period, providers began to collect their data, commission studies and hire evaluators to prove the superiority of their goods and services. The focus on learning outcomes has helped fuel a "marketplace of ideas" (Lubienski, 2019), with intermediaries centrally located between the data producers/sellers and users/buyers. Today's marketplace of policy ideas instead closely resembles a cacophonous bazaar – one crowded with knowledge producers and brokers who bring rather diverse interests and approaches to policy. These actors first translate data into evidence, transform said evidence into policy recommendations and then work to persuade governments to adopt those recommendations. Nonetheless, policy recommendations sometimes precede the data, which is subsequently mobilised to legitimate pre-established policy options. Several of us authors suggested using the term "governance by numbers 2.0" (Steiner-Khamsi, Martens & Ydesen, 2024) to capture the shift from data production to data uptake in today's era of information and policy overload. Relatedly, other scholars analysing controversial policy domains such as privatisation–particularly in the contect of social media–have identified phenomena such as "information pollution" and "information silos" (Lubienski et al., 2024).

The call for data for accountability also curiously resonates with those at the other end of the governance spectrum: students. This call translates into self-directed learning and continuous self-assessment. Students buy tests on apps and digital platforms to pursue self-directed learning and self-assessment, even in countries that resisted a rollout of standardised student assessments. Continuously improving one's performance by means of feedback has become an accepted norm, even if the feedback received takes the form of test results. Teachers and schools also have an embarrassment of riches to choose from for their tests, such as those developed by Mindsteps Inc., which are sold for several grade levels and subjects. The adverse outcomes of accountability measures for students, teachers and schools, however, are not to be underestimated (Levatino et al., 2024; Jabbar et al., 2020).

2. OUR COMPLEMENTARY APPROACH TO UNDERSTANDING THE DATA-FOR-ACCOUNTABILITY PHENOMENON

Our team of five co-authors covers a range of interpretive frameworks that complement each other. This allows us to elucidate aspects of the data-for-accountability phenomenon that would have been overlooked had a single author written this book. For some of us authors, the phenomenon manifests as part of a continual cultural evolution, with quantification gaining ground as part of a set of beliefs about rational individual "actorhood" and now morphing into postliberal forms of thought in which data are depicted as transcending individual decisions (Furuta et al., 2023).

Other authors see intergovernmental organisations as norm-setters (Finnemore, 1993). They argue that the OECD's PISA, for example, does more than just compare 81 education systems; it also makes citizens think like PISA, ultimately leading them to believe that regular testing implies effective learning. They contend that different actors repurpose tests in idiosyncratic ways. Reframing education as a form of human capital was both a blessing and a curse for the field, as it brought a windfall of funds from intergovernmental organisations that had previously been concerned only with "hard" aid, such as building bridges, paving roads, and improving infrastructure. The entry of economists into this research area, however, also meant a "rationalisation" of education in a dual sense: how to cut public expenditures and increase revenues on the one hand and how to calculate the outcomes of investments in the education sector on the other. The gradual increase of intergovernmental

organisations working in the education sector – from a handful after World War II to 30 today (see Chapter 6 by Martens) – implies a reframing of education in terms of economic development (the World Bank), economic, cultural and social development (the OECD), and labour organisation (the International Labour Organization (ILO)).

Other authors of this book emphasise power dynamics by highlighting inequalities in administrative capacity when it comes to data collection, drawing attention to the winners and losers of data-for-accountability reform. The dominance of Anglo-Saxon perspectives in shaping the types and uses of data collection cannot be overlooked. Education statistics remain underdeveloped and are rarely utilised in many regions, including some OECD countries. Furthermore, the assumption that a continuous accumulation of data on specific thematic priorities exists is misguided (Shephard & Delprato, 2024). Some priorities have gained prominence (e.g., testing students' learning outcomes), while others have lost traction (e.g., labour market data for manpower planning, which was actively pursued in the 1970s). Certain priorities are only mentioned symbolically, with few indicators and even less funding attached to them (e.g., lifelong learning since the 1990s). Other priorities are statistically underreported (e.g., the proportion of underqualified teachers) or omitted due to "missing data" (e.g., the number of low-fee private schools, which are frequently not formally registered in governmental records). Data are politically charged and used. A well-documented example of this is the use of standardised tests to weaken teacher unions and empower providers, many of whom are profit-maximising businesses or charter schools that operate schools with underqualified personnel or teachers with non-traditional credentials (Weiner, 2013; Jabbar et al., 2020; Marianno et al., 2023).

Finally, some of us view the provision of public education, for better or worse, as part of broader public administration reforms. As uncomfortable as it may be to acknowledge, ministers of education, whether voluntarily or involuntarily, follow the lead of ministers of finance. The latter serve as counterparts of the two intergovernmental organisations that pioneered data-for-accountability or governance by numbers: the OECD and the World Bank. That said, our explanations are not mutually exclusive. We each attribute different explanatory powers to the individual lines of argument summarised above.

3. SHAPESHIFTING AND COALITION BUILDING OVER A BOUNDARY OBJECT

The assertion that different actors have bought into the data-for-accountability reform for different reasons is not just academic sophistry – it is indicative of social change. A few explanations may be necessary to support such a bold statement.

We share the observation that periods of reform intensity (such as the neoliberal reform of 1992–2008; see Chapter 2 by Bromley) are followed by periods of reform idleness, consolidation or equilibrium. Periods of equilibrium are periodically punctuated or fractured, opening up windows for change. Whether change is realised depends, among other things, on whether stakeholders and other groups in positions of power buy into this change. Coalition-building is indispensable for change. While coalition partners do not need to share the same values and beliefs, they need to see the benefits of change for themselves and, ideally, also for others. For this reason, change happens when different groups buy into a reform for their own reasons.

This brief sketch summarizes and simplifies how and under what conditions social change occurs, and why sensemaking matters. Dissecting the sequence of the change process, as summarised above, opens up a myriad of research questions. Shocks, events or crises are not triggers for change in and of themselves unless they are perceived and framed as problems that deserve political responses. This explains why the theories on the policy process are eager to explore agenda setting and the timing or the period of greatest receptiveness for change (see Shivakoti & Howlett, 2022). Similarly, a significant portion of the research addresses factors that promote coalition-building. Comparative education policy studies, for example, contend that "externalisation" (i.e., references to international standards or lesson-drawing efforts from other countries or sectors) has a salutary effect on coalition-building because these quasi-external reference points come across as neutral, decontextualised and depoliticised (Steiner-Khamsi, 2025). Finally, this book fills a gap in the research on governance by numbers by examining the different rationales, pathways and manifestations of the reform among various actors, levels and time periods.

Bromley (Chapter 2) applies a macroscopic, institutionalist framework to examine policies from 1980 to 2020 that emphasise accountability, autonomy and responsibility. Her analyses use the most comprehensive

database developed to date: the World Education Reform Database (WERD). As principal investigators of the WERD research programme, Bromley and Kijima compiled over 14,375 policy documents from 216 countries and territories, most of which were produced from 1970 to 2021.

Bromley examines the shifting global discourse on education reform, showing that as accountability has become a dominant focus of reforms, the emphasis on responsibility and autonomy has waned. Bromley's chapter situates this transition within the broader trajectory of neoliberal reform, which gained prominence in the 1990s by promoting market-driven education policies centred on decentralisation, privatisation and deregulation. Neoliberal changes initially framed school governance under the paradigm of "school-autonomy-with-accountability", in which increased institutional independence was counterbalanced by extensive accountability measures. However, the chapter reveals a marked discursive shift in the post-neoliberal era: accountability mechanisms, particularly data-driven assessments, continue to expand as explicit discussions of responsibility and autonomy decline.

Drawing on empirical evidence from the WERD, this chapter employs content analysis to track these trends. The data show that since the mid-2000s, mentions of "responsibility" and "autonomy" in education policies have declined significantly, whereas references to "accountability" have steadily increased. Related terms such as "data," "assessment," "evaluation" and "monitoring" have also gained prominence, underscoring the centrality of oversight mechanisms. These patterns suggest a transformation in the global education governance model, moving from systems that simultaneously emphasised autonomy, responsibility and accountability towards systems characterised by extensive data collection and external scrutiny, often without a clear assignment of responsibility to or autonomy for specific actors.

Chapter 2 explores the conceptual distinctions between accountability, responsibility and autonomy to interpret these findings. Accountability, as commonly defined, entails being answerable to an external authority through compliance, reporting and performance measurement. Responsibility, by contrast, is an internally driven obligation tied to ethical and professional commitments. Autonomy refers to the capacity for self-governance, which, in principle, enables responsible decision-making. While neoliberal education policies once framed autonomy as a companion to accountability, the contemporary governance model

appears to decouple these concepts, treating accountability as an end in and of itself.

This shift aligns with broader transformations in governance and public administration, particularly the rise of datafication—the increasing reliance on digital technologies, big data and algorithmic decision-making. The chapter discusses how the expansion of data-centric accountability practices may mark a transition from traditional neoliberal governance to a form of "surveillance capitalism," in which governance is increasingly mediated by predictive analytics and automated monitoring. These developments raise critical questions about the implications of accountability without responsibility: if accountability mechanisms focus solely on data-reporting without assigning clear responsibility for educational outcomes (or assigning responsibility to loose collectives like "the public"), how can education systems ensure meaningful improvements?

Bromley concludes by calling for a deeper investigation into how education policies distribute responsibility, identify who benefits from data-driven accountability structures, and examine how educators, students, and institutions are reconstituted as actors within these evolving governance models. By situating these trends within both historical and contemporary frameworks, Bromley's chapter contributes to ongoing debates about the role of accountability in education policy and governance.

Verger's mesolevel analysis (Chapter 3) compares three cases (Catalonia/Spain, Chile, Norway) that adopted the global school-autonomy-with-accountability reform both on paper and, more selectively, in practice, focusing on performance-based assessment (PBA). These three pathways illustrate the following: (1) a market-driven accountability model marked by increasing scope and consequentiality (Chile), (2) a quality assurance approach primarily rooted in professional control and trust though manifested differently across local contexts (Norway), and (3) the emergence of PBA within a neobureaucratic system with low levels of institutionalisation (Catalonia/Spain). Verger draws on the concept of instrumentation – that is, the process through which educational authorities define and pursue policy goals by selecting and enacting specific instruments – from a comparative policy perspective. He views this phenomenon as being shaped by processes of problematisation, institutional path dependency and policy feedback.

In Chile, PBA emerged within a marketised education system, initially to support school choice and competition under the country's military regime. Over time, Chile's census assessment, Sistema de Medición de la Calidad de la Educación, took on regulatory and equity functions,

resulting in increased accountability pressure on schools. Despite efforts to soften its high-stakes nature, recalibrations have primarily expanded the system. The school improvement industry has reinforced PBA as a governance tool. Norway adopted PBA in response to a "learning crisis" revealed by PISA. Resistance led to recalibrations that strengthened professional control, which made national tests more formative. However, municipalities and the local media in Norway have intensified accountability pressures. Norway's case illustrates the challenge of balancing equity-driven education with increasing performance monitoring. Catalonia followed a fragmented trajectory, implementing PBA within a neobureaucratic framework with weak institutional capacity. Large-scale assessments were designed for formative purposes but became instruments for school inspectors and other educational authorities to exert supervisory control – a practice legitimised and expanded by recent school improvement mandates.

By emphasising how various dimensions of instrumentation shape policy pathways, the chapter demonstrates that PBA policies do not adhere to a linear or standardised adoption process. Instead, these pathways depend on pre-existing institutional arrangements, governance traditions and the individual agency of policy actors, including key instrument constituencies. Additionally, feedback mechanisms reinforce or recalibrate policy frameworks over time, often resulting in unintended consequences and varied policy patterns. While Chile illustrates a path of accumulation and intensification, Norway reflects a reflexive recalibration approach and Catalonia exemplifies a stalled and weakly institutionalised model. These findings deepen the broader understanding of how global education reforms are translated into diverse national and subnational settings, challenging deterministic views on policy diffusion and implementation.

Steiner-Khamsi (Chapter 4) combines the two levels of analysis mentioned earlier: the macrolevel unit of analysis (global trends over 60 years) and mesolevel examinations of how governmental institutions adapt and translate these global trends in path-dependent or idiosyncratic ways. The disaggregation of institutional responses and differentiation between two institutional trajectories (ministries of finance versus ministries of education) is a unique twist in Steiner-Khamsi's meso-level analysis. These two types of ministries are unequal in terms of decision-making authority, their proximity to powerful global actors and their temporal ordering in which innovation (good or bad) is diffused within government. Typically, the line ministries, such as the ministries

of education, depend on and therefore follow the lead of ministries of finance. Acting as the national counterparts for public administration reforms, ministries of finance are directly exposed and see themselves as implementers of OECD and World Bank reform agendas.

Drawing inspiration from sociological systems theory, which situates systems in their broader environments or institutional contexts, Steiner-Khamsi pays close attention to how the two types of ministries translate the objects and objectives of datafication into the education sector and their system logic. The use of data for accountability purposes has undergone significant transformations over the last four public administration reforms, which have been amply discussed in public policy studies. The role of the state in public policy has arguably changed with every reform. The redistributive and equity-oriented welfare state (1960s/70s) was preoccupied with compliance and enforced state-based accountability. The subsequent reform, the neoliberal entrepreneurial state (1980s/1990s), attacked the strong and large state apparatus and asked for its replacement using market-based accountability. To mitigate the ensuing inequality and erosion of quality resulting from rampant privatisation and outsourcing, the interventionist state (2000s) introduced a Third Way approach to governance, which encouraged public–private partnerships on the one hand and, on the other hand, introduced standards-based accountability.

In education, this entailed developing standardised curricula, student assessments and accreditation policies. Currently, we are in the period of the engaged state, which aims to reduce the state apparatus further and shift towards greater citizen engagement and social accountability. As with previous reforms, the OECD, the World Bank and their counterparts at the national level – the ministries of finance – advance and fund the current e-government reform (labelled in some countries as "new public governance"). The two global actors demand that governments share data publicly to strengthen social, public or demand-side accountability. How did ministries of education translate these public administration reforms into education? Drawing on global trend analyses and case studies, Steiner-Khamsi shows how the object of datafication was repurposed over time and continuously expanded in scope, first from the education sector (welfare state), the school (entrepreneurial state), the class/teacher (interventionist state) and, most recently, to the student (engaged state).

Kijima (Chapter 5) analyses the trends and patterns of reforms associated with assessments. She categorises assessment reforms using the

comprehensive compilation of policy documents assembled and made publicly available in the third version of the World Education Reform Database. Chapter 5 explores the changes in the patterns of assessment reforms that are shaped by neoliberal policies in education and categorises assessment reforms into three levels: (i) reforms that focus on public accountability, (ii) systems and mechanisms for assessment reforms, and (iii) assessments as a measurement of learning that involve students and teachers. The overall trends across these three distinct yet intertwined dimensions indicate an overall increase in assessment reforms over a period of 60 years. In addition to conducting trend analyses at the global level, this chapter brings institutional and individual levels into focus.

A granular analysis of assessment reforms at the institutional level uncovers the three key institutional mechanisms of accountability: monitoring and evaluation, institutional capacity-building, and accreditation and quality assurance. This study reveals that the category of institutional mechanisms constitutes the largest proportion of assessment-related reforms. This study's key finding is that assessment reforms are often intended to be functional: they reference the establishment of monitoring and evaluation systems or system-wide information systems, such as education management information systems (EMIS). Reforms associated with monitoring and evaluation as well as quality assurance and accreditation have recently been on the decline. Some possible explanations for these findings are presented in this chapter.

Kijima's chapter, unlike the others in this book, introduces a third level of analysis: the micro level, which focuses on reforms related to students and teachers. The chapter reveals an increase in the proportion of assessment reforms associated with students and teachers. An analysis of these reforms shows that the nature, content and intent of assessment reforms have shifted over time to involve greater standard setting and oversight of student competency and teacher performance.

Martens (Chapter 6) invites us to examine global actors and their continually evolving tools of global governance from an international relations perspective. Without the ability to impose binding laws or coerce states, how do intergovernmental organisations (IOs), such as the OECD, UNICEF, UNESCO, and World Bank, legitimise their existence, and how do they influence governments? This chapter examines how education has become a field crowded by IOs. Since WWII, more and more IOs have taken up activities in education policy. Today, around 30 IOs can be identified as Education IOs. For some IOs, education has always been a relevant topic due to their mandate, while many others have only

taken up the topic during their existence and for different reasons. In total, 30 IOs, ranging from global to regional and from cultural to economic, are present in the field of education today.

The chapter examines the mechanisms by which these IOs contribute to global education scripts and how they disseminate their data, beliefs and recommendations with which they can hold nation-states accountable. With very few exceptions, all IOs contribute to the global script of educational accountability. In particular, the IOs contribute to the global diffusion and transfer of policy ideas and norms through policy briefs and documents. At the same time, only a few have "hard" mechanisms of funding and conventions. Martens' findings suggest that global actors can apply the full range of mechanisms to influence national policy. At the same time, many regional IOs have only a limited spectrum at their disposal when disseminating policies. In line with the neo-institutionalist framing of world society, Martens conceives of IOs as organisations that move beyond the national agendas of their member states and instead foster "accountability from above". Although IOs in education, akin to IOs in other fields, have only limited means of enforcement, they are effective at collecting and providing comparative data. Therefore, IOs are considered a critical element in developing and implementing scripts in education. PISA – the OECD's primary tool for measuring and assessing education – stands out as the global tool for generating and evaluating data.

4. CONCLUSIONS AND AGENDA FOR FURTHER RESEARCH

Our team of five co-authors has analysed data-for-accountability in education from five distinct yet complementary perspectives. Each of us authors assigns varying levels of importance to the different levels of analysis (macro, meso and micro), interpretive frameworks (neo-institutionalist theory, political economy, historical institutionalism and sociological systems theory) and methods of enquiry (e.g., case study methodology, quantitative analyses and sequential analyses). These nuances are significant and are detailed in the book's chapters. Nevertheless, we all believe that the panoptical view is essential for fully understanding how various actors -- including students, teachers, parents, schools (public/private), governments (district, sub-national and national/federal), international organisations, associations/unions, the private sector/philanthropies as well as donors -- have repurposed data

for accountability over time. Depending on their relative power with respect to other actors, education actors have changed how they use data for accountability in alignment with political, social and economic constellations across various institutions. We show how the macro-, meso-, and micro-levels intersect, thereby complicating any attempt to separate them as distinct analytical categories. The fluidity of these levels and their intersectionality have explanatory power in and of themselves. For example, politicians use developments in other countries as leverage to either justify a reform or alleviate reform pressure in their own countries. Until recently, such politicians did so by pretending that globalisation was an inevitable external force steering educational reform in specific directions.

Given the rise of nationalist populist movements in different parts of the world and the accompanying scepticism towards global institutions and initiatives, many new avenues of research are likely to open. There is ample room for speculation about how new actors will, yet again, repurpose data to serve their agendas or, alternatively, disregard scientific data and evidence in their policy frameworks. At this stage, it is too early to offer definitive answers to this emerging phenomenon. What we do have, however, are research questions that guide future enquiry. Specifically, we make three propositions for an agenda for further research on these new developments.

First, large-scale student assessments have experienced exponential growth in all dimensions over the past 25 years across more and more countries, more and more subjects, more and more types of assessment, and at each level of schooling (Kamens & McNeely, 2010; Kijima & Lipscy, 2024; Verger et al., 2019a). Has this boom reached a point of saturation, burnout or even reversal? How are governments responding to distrust or public doubt about the utility of such international comparative studies? We appear to be at a crossroads in which we can choose to either scale back testing or intensify it and make it more "consequential." Decisions such as linking test results to financial incentives or reputational consequences often follow pendular dynamics and are heavily influenced by political shifts and feedback loops from both policy and practice.

Second, will performance measurement or governance by numbers, prominent since the 1990s, be replaced with a direct feedback system made possible by advanced computational systems? What does this entail for monitoring and evaluation? In 2024, X moved to replacing fact-checking with its User Notes feature. Users, not providers or funders, are

responsible for regulating the site. Meta followed suit in 2025. What are the implications for education as citizen engagement and social accountability are rigorously translated into the education system? Can such models scale and spread internationally? Will parents and communities become central actors in quality assurance practices, thereby potentially displacing the role of educational authorities and the state?

Finally, IOs have been active advisors and, in some cases, funders of governments promoting global trade and the transnational flow of ideas, goods, services, finance and school reforms. Some countries have benefited from globalisation more than others. How do the mechanisms of global governance change in the post-globalisation world order? The current global coordination mechanism propels the use of research evidence for policy, planning and practice (Baek, 2024). More specifically, learning and other types of educational data are increasingly viewed as lingua francas that transcend national boundaries and allow global actors to monitor progress towards international agreements, such as the Sustainable Development Goals. The rapid advancement of generative AI provides an opportunity for scholars and policymakers to dig deeper into determining the kind of international data used in educational debates in the future. Will such debates leverage the big data created by billions of people who continuously produce data merely by using the internet? Given the uneven outcomes of globalisation, it is crucial to remember that some people and some countries have more access to the internet than others and thereby share, voluntarily or involuntarily, their data. In what ways can AI bring greater efficiency to the education system, thereby affecting the way students learn, teachers evaluate and governments manage such a large corpus of data?

2. Accountability without responsibility? The era of datafication in education reform

Patricia Bromley

During the 1990s, a neoliberal reform movement swept across the globe, as outlined in the introductory chapter of this volume. These reforms introduced market-based policies emphasising deregulation, decentralisation and privatisation (Adamson et al., 2016). Although these policies shifted authority, decision-making and responsibility away from centralised state structures, they were swiftly accompanied by the development of extensive accountability systems, thereby reintroducing government oversight in a novel form. This package of reforms is frequently referred to as "school-autonomy-with-accountability" (SAWA) in the education sector (e.g., Crawford, 2001; Arcia et al., 2011; Verger et al., 2019c, 2019a). For a time, neoliberalism served as the dominant global paradigm for structuring state and society, with some even declaring neoliberalism to be the "end of history" (Fukuyama, 1992). In recent years, however, the hegemony of neoliberal ideology has visibly diminished. Notably, even economists at the World Bank and the IMF—institutions that once forcefully championed neoliberal reforms—have acknowledged that the benefits of neoliberalism may have been "oversold" (Ostry et al., 2016). As confidence in neoliberalism wanes, critical questions concerning the evolving focus of education reforms arise.

This chapter draws on a global database of education reforms to examine whether the dominant global education script is undergoing a transformation, paying particular attention to trends related to responsibility, accountability and autonomy. The content analysis of reform keywords conducted for this chapter reveals a decline in the language of autonomy and responsibility since the apex of neoliberal influence. However, education policies have increasingly emphasised accountability practices.

After clarifying the distinctions between the concepts of accountability, autonomy and responsibility, this chapter presents empirical evidence of evolving trends in global education policy discourse. It subsequently explores the implications of these changes, offers potential explanatory frameworks and concludes by identifying promising directions for future research.

This chapter analyses national policy changes recorded in the World Education Reform Database version 3 (WERD) (Bromley et al., 2023) to illustrate changes in the discourse about global education governance.[1] The database encompasses reforms from 204 countries and territories, thereby ensuring broad geographic representation.[2] Reforms—used interchangeably here with policy changes—are drawn from reports produced for international organisations or authored by, with or for international education experts.[3] International organisations generate a substantial number of such reports, often providing country-specific snapshots of education systems. For instance, the OECD's Education Policy Outlook series monitors the evolution of policy priorities and

[1] Additional details on WERD are available in a methodology document published online at https://werd.stanford.edu/. In this database, the term "reform" is operationalised as intentional ("planned") change ("non-routine") occurring at an administrative level above an individual school (part of a "system"), as distinguished from unplanned, non-systemic, non-formal education fads or small-scale initiatives. Research assistants are trained in this definition and manually code reforms by reading selected documents and collecting the start years, titles, and brief descriptions of each policy. The intraclass correlation for coding reforms using these criteria is 0.85, which is considered "excellent" by common standards for inter-rater reliability (Cicchetti, 1994).

[2] No single country makes up more than three per cent of the sample of reforms. Appendix A shows the distribution of reforms used in this study, thereby illustrating its wide global coverage.

[3] The scope and scale of changes within the concept of reform used by WERD runs from incremental first-order policy change to second-order strategic change to third-order paradigm shifts (Hall, 1993). In addition, reforms come in a variety of forms, such as education legislation (e.g., acts, laws and decrees), key policy shifts (e.g., new national strategic plans, initiatives, guidelines and standards) or the creation of new educational organisational structures that are part of the government (e.g., new research boards, ministries and departments).

developments across education sectors, primarily within OECD countries, to foster a comparative understanding and inform policy improvement (OECD, 2025). Similarly, the World Bank's Systems Approach for Better Education Results initiative, launched in 2011, aims to produce comparative education data and policy insights (World Bank, 2025b). Reports prepared for or by international organisations thus offer valuable insights into globally legitimated education policies (Bromley et al., 2023). Furthermore, while these reports pertain to specific countries, they are best understood as reflections of national engagement with global education discourses rather than as purely domestic narratives.[4]

The production of education reports constitutes a critical mechanism through which international organisations act as "teachers of norms" and exert soft-power influence over national education systems (Finnemore, 1993; McNeely, 1995; Niemann & Martens, 2018; Elfert & Ydesen, 2023). For example, Verger et al. (2019c) demonstrate that despite limited legal and financial coercive powers and a relatively narrow membership base, the OECD has played a central role in promoting the SAWA model across the globe. The OECD's influence stems not from direct enforcement but from its capacity to generate policy ideas and shape evaluation frameworks. Thus, reports produced by organisations such as the OECD serve as pivotal vehicles for disseminating preferred education models. Chapter 6 of this volume provides a more detailed examination of the role of international organisations in shaping education reforms worldwide.

Formal reports and policies articulate legitimised organisational practices and educational aspirations, often establishing normative frameworks. Nevertheless, a well-documented gap frequently exists between the discourse presented in reports and actual practices on the ground—a

[4] The reforms in version 3 of the database were drawn from 1,389 reports across nine sources: 1. World Bank Project Appraisal Documents (413 reports), 2. World Bank Systems Approach for Better Education (176 reports), 3. UNESCO's World Data on Education (252 reports), 4. the International Encyclopedia of Education (279 reports), 5. the OECD's Review of National Policies for Education (50 reports), 6. the OECD's Education Policy Outlook (39 reports) and the OECD's Reform Finder Database (38 reports), 7. the SHARE Survey (Survey of Health, Ageing and Retirement in Europe) (16 reports), 8. the World Education Encyclopedia (126 reports) and 9. OECD's Education Reforms Finder database.

phenomenon termed "decoupling" in sociological literature (Meyer & Rowan, 1977; Bromley & Powell, 2012). Decoupling does not result in formal structures being rendered inconsequential. Discursive trends reflect and shape legitimate actor identities and organisational goals, thereby influencing how actors interpret and respond to their external environments (Czarniawska-Joerges & Sevón, 1996; Sahlin & Wedlin, 2008).

Recognising that the processes generating policy content and those driving implementation are distinct and yield different consequences is more accurate than viewing formal policy as merely symbolic. At times, decoupling between policy and practice serves a strategic purpose, and it can occur for many intentional and unintentional reasons (e.g., a lack of resources, misperception, translation to local circumstances or strategic symbolism). Thus, determining when, where and why certain contexts exhibit greater or lesser degrees of decoupling from the global policy script is a critical area of enquiry. For instance, a study on SAWA adoption in Switzerland found that only selected elements were transferred and adapted: while standardised testing and professional school management became institutionalised, performance-based pay was abandoned and external evaluations were largely maintained in a symbolic form (Steiner-Khamsi et al., 2024b).

1. CONCEPTIONS OF ACCOUNTABILITY, AUTONOMY AND RESPONSIBILITY

Accountability, autonomy and responsibility were central mantras of education reform during the neoliberal era. The theory that underpins the neoliberal advocacy of these trends is rooted in economic conceptions of rational action. The theoretical foundation underpinning neoliberal advocacy for these trends is rooted in economic conceptions of rational action. According to this theory, transferring authority to actors closer to schools would ensure more effective and efficient management when coupled with the implementation of systems to monitor progress. While some authority shifted to subnational units or local school districts, it often devolved beyond government systems entirely and into the domain of private actors. As Keddie (2015) notes, "Responsibility for school improvement in this climate has shifted from a state to a non-state matter. While, to be sure, state intervention to support school improvement still exists, the responsibility for school improvement is seen as the domain of the school, local community, family, or individual" (p. 4).

Standard definitions emphasise distinctions among actors who are obligated to achieve outcomes (responsibility), who possess decision-making authority (autonomy) and who must demonstrate compliance or report on activities (accountability).

According to a World Bank report, accountability involves "being answerable for one's actions. In school management accountability [includes]: (i) the act of compliance with the rules and regulations of school governance; (ii) reporting to those with oversight authority over the school and (iii) linking rewards and sanctions to expected results" (Arcia et al., 2011, p. 2). Lindquist and Llewellyn (2003) similarly conceptualised accountability as an external governance mechanism requiring detailed reporting, assessment and justification to an external authority or audience. Accountability is inherently relational and premised on the notion that individuals or organisations must answer to others for their actions and decisions. Olsen (2015) described accountability as "a principle for organising the relations between rulers and ruled" (p. 425). Quantified accountability measures rose to prominence during the neoliberal era. Consistent with market-based ideologies, these measures aimed to render elements of schooling visible for external scrutiny via comparison and competition (Grek, 2009). Policy changes promoting standardised testing, performance metrics and extensive statistical reporting (e.g., on enrolment, completion rates and teacher credentials) became commonplace.

In contrast, autonomy refers to actors' self-directedness and internal capabilities. The World Bank defines autonomy as "a form of school management in which schools are given decision-making authority over their operations" (Arcia et al., 2011, p. 2). Decision-making is fundamental to autonomy, and Western philosophical and legal traditions have linked the capability for autonomous decision-making to the allocation of responsibility (Miller, 2001; Hart, 2008). Although autonomous actors may not always behave responsibly, autonomy is a necessary precondition for assigning responsibility for outcomes. Autonomy is intended to enhance actors' willingness and ability to assume responsibility for achieving educational goals. Bivins (2006) wrote:

> In addition to responsible actors being imbued with the ability and the freedom to make self-regulating decisions, they are also able to motivate (free of outside pressure) their own responsive adjustments to situations in which their decisions have had an impact. This is what separates them from accountable actors, who must rely on external oversight for motivation to respond and adjust. (p. 23)

The distinction between accountability and responsibility is well established in organisational and management theories (Bovens, 1998).[5] Scholars have described how organisations and governments are constructed as social actors who can be accountable and/or responsible to stakeholders. Brunsson et al. (2022) explained this as follows:

> [A]ccountability [is] a common concept in the organization and accounting literature that should not be confused with responsibility (Bexell, 2005; Bovens, 2010; Heimstädt & Dobusch, 2020; Messner, 2009; Roberts, 1991; Roberts & Scapens, 1985). Accountability can be defined as an expectation to provide an account of one's actions after the fact and before some sort of public forum (Bovens, 2010; van Gunsteren, 1976) … But whether the account will actually make the person responsible is far from certain. And responsibility may well be attributed to a person who is not accountable and has not rendered an account. In fact, there may be responsible persons, even if no one is accountable. (p. 3)

Understanding the distinctiveness of autonomy, responsibility and accountability is crucial for analysing governance in education. Accountability is externally imposed and anchored in institutional authority or public expectation, while responsibility is internally driven and rooted in actors' obligations to personal or collective principles. Bivins (2006, p. 23) noted, "A responsible being is a being who can make choices according to his or her own insights. He or she is not under the control of others." On the other hand, the accountable actor is "held to external oversight, regulation and mechanisms of punishment aimed to externally motivate responsive adjustment in order to maintain adherence with appropriate moral standards of action" (Bowie, 1982, pp. 95–96, as cited in Bivins, 2006). Responsibility encompasses duties that extend beyond the need to justify specific actions. That neither autonomy nor accountability guarantees responsibility is a critical distinction often overlooked in contemporary discussions Figure 2.1 illustrates these conceptual relationships: autonomy is necessary for responsibility,

[5] An exception to this can be found in the concept of individual responsibility among education professionals discussed by Glatter (2012, p. 562), who observed "a conceptual distinction between accountability and the general responsibility that any worker feels, or should feel, towards those affected by his work." Glatter's definition of accountability was reported as "the duty to render account of work performed to a body that has authority to modify that performance, by the use of sanctions or reward" (2012, p. 562).

but autonomous actors may act irresponsibly. Further, accountability and autonomy may coexist without responsibility, and "irresponsible accountability" may emerge, as exemplified by "gaming" accountability systems (e.g., manipulating university rankings) or corporate "greenwashing" practices.

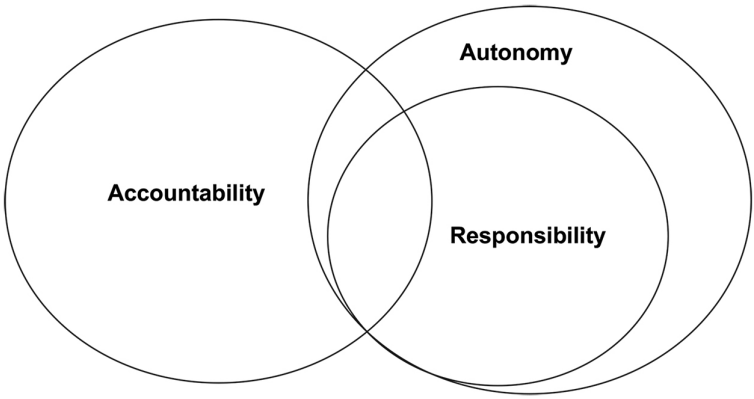

Figure 2.1 *Conceptual diagram of responsibility, autonomy and accountability*

2. TRENDS IN POLICIES USING THE TERMS ACCOUNTABILITY, AUTONOMY OR RESPONSIBILITY

Figure 2.2 illustrates trends in the terms "accountability," "autonomy" and "responsibility" in education policy changes reported since 1980 in the set of reforms in WERD.[6] Two trends are most notable. First,

[6] These keyword searches were conducted using Python, which enabled us to count the number of times a word was mentioned in each reform and to develop code to specify necessary permutations and cleaning rules for each term. Drawing on the literature and expertise in education, the author and a research assistant generated an extensive list of relevant terms. To analyse the data, first we stemmed terms when needed so that they included all relevant permutations, which is, in accordance with convention, indicated

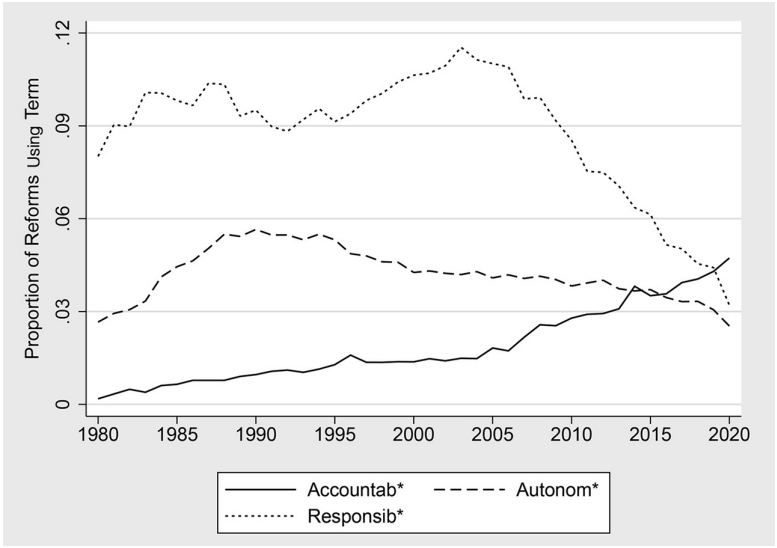

Figure 2.2 *Proportion of policies using the terms "accountability," "autonomy" and "responsibility"*

since 1980, there has been steady growth in the percentage of reported reforms each year that mention accountability. For example, a reform reported from Kenya in 2005—discussed in Box 2.1—exemplifies an accountability-driven policy change. Second, in contrast, references to

with a "*" at the end of the stem. For example, "monitor*" includes "monitors", "monitored", and "monitoring"; "autonom*" includes "autonomy" and "autonomous", and "responsib*", which includes "responsible", "responsibility", and "responsibilities". Second, to clean the data for each keyword, the author and a research assistant read through the returned results and revised the code to remove unwanted results. For example, in early iterations of this work, the term "ICT", which we intended to capture "information and communication technology", returned results for the words "conflict", "district", "jurisdictional", "dictionaries", and "Victoria". Third, we collapsed the reform-level data into an annual dataset to obtain the proportion of reforms per year rather than mention each term. Note that additional terms, such as "metric" and "benchmark", were analysed and omitted from the results because they appeared too infrequently to reveal any trends.

autonomy and responsibility have declined since peaking at the height of the neoliberal era. As shown in the examples from Mexico, Senegal and Sweden in Box 2.2, many responsibility- and autonomy-related reforms emphasised decentralisation, which is consistent with observations in prior research (Keddie, 2015). The percentages reported here reflect this overall trajectory, capturing a wave of expansion during the 1990s and the early 2000s.

BOX 2.1 EXAMPLE OF ACCOUNTABILITY REFORM

Kenya (2005): The key elements of the future strategy will include an extension of efforts to increase the dissemination of information at the secondary and tertiary levels, even more accountability to the public, and continuous capacity-building and independent monitoring of activities. For example, at the secondary level, the measures will include improved guidelines for the management of resources; capacity building in financial management and procurement; public expenditure tracking; the posting of national examination results and Form I intake on the website; revising the code of ethics for teachers and other enforcement mechanisms; and extending social accountability (with increased community involvement) and the use of risk-based internal audit.

BOX 2.2 EXAMPLES OF REFORM MENTIONING RESPONSIBILITY

Mexico (1992) "National Agreement for the Modernisation of Basic Education": As part of the wider reforms of school education in the National Agreement for the Modernisation of Basic Education, **responsibility** for federal public normal schools was transferred to the states, although **responsibility** for specifying curricula, staff conditions and the majority of institutional funding remained at the federal level.

Senegal (1993) Deconcentration policies: **Responsibility** for administration and staffing has been devolved to the regional Inspectorates for Education.

Sweden Education Act (Skolllagen, SFS) (1985): **Responsibility** for school management is assumed by the municipalities, which are obliged to provide a number of basic services and ensure that Swedish schools uphold equivalent standards. The municipalities are granted extensive **autonomy** in administering them.

The pattern observed for the first several decades aligns with the trends described by Verger et al. (2019a) in which initial attention to autonomy was followed by an intensified focus on accountability.[7] However, the trend has shifted since the mid-2000s. Mentions of "responsibility" in reforms have declined sharply—falling from a consistent rate of approximately 10 per cent of reforms to less than 5 per cent. Similarly, references to autonomy have also diminished. In contrast, mentions of "accountability" have continued to increase, rising from minimal occurrences in the 1980s to nearly 5 per cent by 2020—surpassing both responsibility and autonomy.

2.1 Further Analyses

While trends in the three central terms provide important insights into the evolution of education policy discourse, further investigation is warranted. First, it is plausible that these trends capture only narrow language uses and fail to reflect broader conceptual shifts that may be articulated through evolving semantic forms. Second, examining only general term frequencies overlooks potential changes in specific practices or dimensions associated with broader concepts. To address these concerns, Figures 2.3 and 2.4 analyse additional keywords linked to accountability, autonomy and responsibility.

Terms conceptually related to autonomy and responsibility, such as "decision" and "choice", display similarly stagnant trends over time. In contrast, terms linked to accountability, such as "data," "assessment," "evaluat*," and "monitor", show marked increases, particularly over the

[7] These global trends mask several variations, such as early and later adopters of such reforms or varied prevalence between countries. Determining when, where, and why countries diverge from the global script is an important area of study, as stated earlier.

past two decades. Chapter 5 (Kijima) of this volume provides a valuable deep dive into the realm of assessment reforms, offering additional insight into these trends. Figures 2.3 and 2.4 indicate that the broader patterns observed in the central concepts are mirrored in related terms, suggesting that the trends are not artefacts of specific word choices. Moreover, many terms associated with accountability emphasise highly quantified and datafied forms of governance. Alternative mechanisms, such as school inspections—while common features of accountability systems (e.g., Ehren et al., 2015)—are not referenced frequently enough in reforms to generate visible trend lines.

This description of reported education reforms reveals a post-neoliberal script that involves a hyperfocus on accountability policies, supporting arguments that we are in an era of information overload that can be described as "governance by numbers 2.0" (Steiner-Khamsi et al., 2024b). Adding to insights about a new era of global educational governance in prior research, our empirical analysis reveals that the post-neoliberal script involves not only a continued ramping up of accountability reforms but also a retrenchment of emphases on responsibility and autonomy.

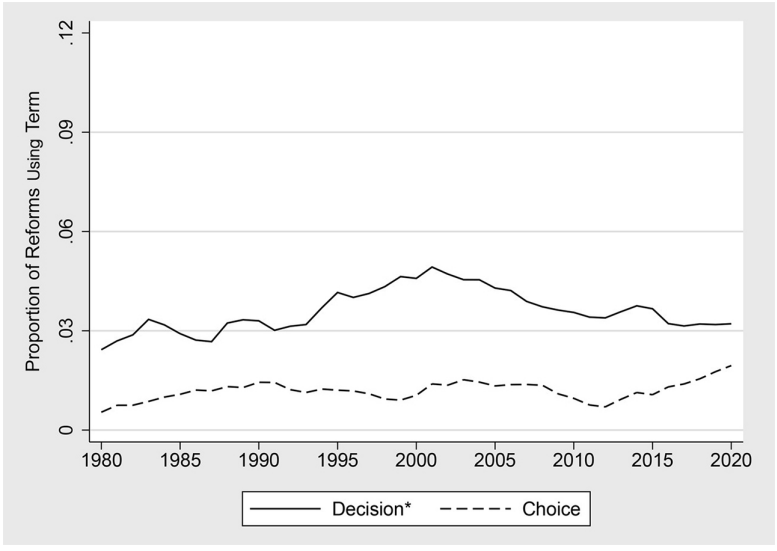

Figure 2.3 *Stagnation of terms related to autonomy and/or responsibility (decision and choice)*

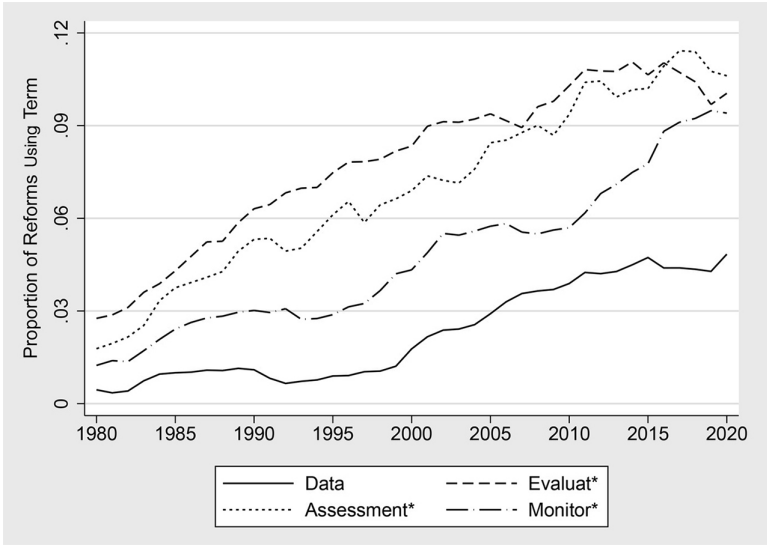

Figure 2.4 *Increase in terms related to accountability (data, assessment, evaluat* and monitor*)*

3. IMPLICATIONS

The conceptual discussion above outlines the meaningful distinctions among accountability, autonomy and responsibility and highlights an important implication of these trends. Accountability does not ensure responsibility. Scholars and practitioners have argued that a healthy and high-functioning governance system requires a balance of accountability and responsibility/autonomy (March & Olsen, 1995; Bivins, 2006; Bogotch, 2014). For example, Lindquist and Llewellyn (2003) contended that successful educational organisations require a balance between formal accountability systems and a culture of professional responsibility (see also Hargreaves et al., 2013, p. 26). Accountability structures provide a necessary framework for ensuring that educational institutions meet societal and governmental expectations; however, they must be sufficiently flexible to allow educators the autonomy to act responsibly when fulfilling their pedagogical and ethical duties.

Professional autonomy and responsibility are essential for achieving important outcomes, especially those that are difficult to observe or quantify or that require ethical decision-making. Reflecting the recognition that both accountability and responsibility/autonomy are vital, many education reforms have sought to introduce these elements simultaneously. Box 2.3 presents examples of reforms from New Zealand and Italy that embody this dual emphasis. Although determining the ideal balance between accountability and responsibility is challenging and likely varies depending on context, the pendulum has evidently swung towards accountability at the expense of responsibility relative to historical levels.

BOX 2.3 POLICY CHANGES EMPHASISING RESPONSIBILITY/AUTONOMY WITH ACCOUNTABILITY

New Zealand (1989): From 1989, the administrative structure of every sector of the education system from early childhood to tertiary level has undergone major reform. Features of the reform include decentralisation of greater **responsibility** to each individual institution in such areas as the setting of goals and aims, finance management, and staff appointments; measures to increase **accountability** of individual institutions both with regard to the local community and the Minister of Education; and new procedures for reviewing the performance of institutions against their agreed goals.

Italy (1997) "Act on School Autonomy": "This approach raises the issue of what will be the role and **responsibility** of the central authority if management is delegated to regions. Obviously, the centre's role will be much more strategic and will provide national frameworks and standards. The act retains state **responsibility** for matters such as the structure of the system, curriculum, general organisation of schooling, and legal status of the personnel. This approach is similar to that adopted in other countries that have moved in the direction of greater devolution, such as Spain. This combination of a core set of strong **responsibilities** of the state with a wide **transfer of management responsibilities** to regions and municipalities fits with the demands of greater **autonomy** at the school level, accepting that a common educational core

determined nationally is retained. Nonetheless, if the state does not have instruments and processes to **monitor** the implementation of such a core in practice, then the law will be no more than rhetoric. Giving the law "teeth" and supporting its intent will require the establishment of a nationwide **system of evaluation**, combined with new mechanisms for school **accountability** and a strong central inspectorate."

The fact that the terms "data" and "assessment" are increasingly prominent words in education reform and are often used to describe elaborate accountability systems is another significant implication of this study's findings. These reforms are linked to and amplified by broader societal trend towards datafication, where large-scale data and information systems are perceived as the most legitimate and effective tools for achieving progress and justice (Sadowski, 2019). Revolutions in computing power, data analytics and artificial intelligence profoundly influence the nature of accountability regimes, transforming how monitoring and evaluation are conducted through new possibilities for data collection and analysis.

Some scholars have argued that these new forms of accountability do not merely extend the neoliberal trend towards an "audit society" in which external processes constitute the categories and boundaries of social actors (Power, 1997). Rather, they represent inflection points that mark a shift from neoliberal capitalism towards surveillance capitalism (Zuboff, 2015, 2019; Power, 2022).

Scholars have identified several core distinctions between earlier neoliberal and datafied accountability models. First, data-driven approaches behave as a form of self-governance that operates outside external scrutiny and oversight (Cohen, 2019). For example, accountants or old-fashioned school inspectors are embedded in professional guidelines and operate under a principle (albeit often tenuous) of independence. However, data scientists and software engineers who create new accountability systems and generate analyses render this myth of accountability via external audits largely irrelevant; rather than verifying compliance on behalf of an authority, data experts design tools that automate assessment processes and broadly disseminate information. As a result, the internal/external and ruler/ruled dualisms of accountability in the audit society have eroded (Stark & Pais, 2020).

Second, the logic underpinning accountability activities has undergone a fundamental shift. Neoliberal models are grounded in myths

of enhancing efficiency and effectiveness through informed decision-making. In contrast, new datafied systems are driven by a deeper logic of surveillance and security (Cohen, 2019). In the datafied ideal, data extraction is continuous and granular, encompassing all possible actions to minimise uncertainty (Ananny & Crawford, 2018; Zuboff, 2022). As Power (2022, p. 13) describes, "Where the old accounting problem was that of informing decision-making under conditions of uncertainty, the dream of accounting in the world of surveillance capitalism is the production of certainty and knowledge of everything." In this context, traditional acts of decision-making become peripheral, supplanted by algorithmic assessments that are envisioned to yield self-evident answers based on vast quantities of data. Under the logic of datafication, traditional accountability mechanisms have limited measurement capacity; thus, they require greater professional autonomy and responsibility to act in the unmeasurable arenas. To make this logic explicit, through full surveillance and measurement, responsibility can be replaced by control. Naturally, ethical and moral questions regarding this datafied approach should be asked.

Third, the constitutive nature of accountability practices may change education. Under neoliberal models, accounting practices and their related features, such as ratings and rankings, are externally created categories that, in part, constitute actors, defining actors' activities and identities. In a datafied society, inductive descriptions of large amounts of information via data gathering can partially constitute actors and identities. For example, analyses of text or video in the classroom can be used to reveal latent groupings of students (e.g., in terms of how they interact with each other, the teacher or engage in class). These data-driven patterns may or may not overlap with externally prescribed categories, such as indicators of language learning or learning differences. While such all-encompassing data analytics seem unlikely in most educational settings, at least until well into the future, the discursive shift discussed in this section indicates that neoliberal ideology is changing.

4. EXPLAINING THE TRENDS

Trends in the education reform discourse of declining discussions of responsibility and autonomy alongside increasing emphases on accountability can be explained in several ways. Reforms, when viewed through a functionalist perspective, are technical interventions aimed at improving the efficiency or effectiveness of educational systems. Thus, reforms

target areas perceived to require enhancement. Accordingly, the declining emphasis on responsibility and autonomy could be interpreted as evidence that these aspects have been satisfactorily addressed. The remaining challenges in this field—and thus the continuing reform efforts—focus primarily on strengthening accountability mechanisms. This diminished attention to responsibility can optimistically be understood to suggest that responsibility is appropriately allocated. Conversely, this could also indicate that the state has privatised responsibility and is no longer able to legislate reforms in this domain. The principal task in either interpretation is to ensure that newly empowered actors fulfil their obligations through accountability structures. Although these arguments may partly hold, they often overlook the political and social complexities of reform processes, particularly the challenges that arise when responsibility is diffusely allocated across a wide range of actors.

Many scholars have contended that neoliberal-era reforms were not merely technical adjustments but deeply political projects aimed at transferring responsibility from the state to individuals as a means of control. Rose, O'Malley and Valverde (2006) described how neoliberal reforms sought to produce "the ends of government by fulfilling themselves rather than being merely obedient" (p. 89). Discussions of neoliberal "responsibilisation" build on Foucault's concept of governmentality, suggesting that individuals are encouraged to embrace active responsibility-taking through promises of autonomy and personal freedom (Rose, 1992, 1999; Barry et al., 1996; Peters, 2002). The neoliberal logic assumes that individual choice and decision—better reflected in local rather than federal governance—yields optimal outcomes, thereby conflating autonomy, decision-making and responsibility.[8] Trnka and Trundle (2014, p. 138) emphasised that a central theme of neoliberalism is "the portrayal of personal choice and autonomy as the means through which responsibility is enacted." The neoliberal "bargain" between the state and individuals fuses responsibility and autonomy, blurring the distinction between them. In this framework, actors' core responsibility lies in exercising autonomy and choice rather than in upholding broader principles such

[8] In practice, autonomy may not be realised despite it being central to neoliberal representations of the individual or organisational "self" (Rose, 1999). This is partly because the term "autonomy" has meant a break with a centralised authority, often obscuring a transfer from political dependencies to dependencies on private actors or market forces (Olsen, 2009).

as justice. Consequently, distinct emphases on responsibility diminish as autonomy becomes the operative marker of responsible behaviour.

Other authors have also described how accountability comes to be equated with responsibility under neoliberal ideologies. Shore and Wright (2011) discussed how neoliberal notions of the responsible actor are replete with auditing, accountability and assessment practices that promote surveillance by oneself and others, ultimately arguing that actors use accountability enactments as evidence to others that they are fulfilling their responsibilities. In these views, engaging in accountability practices is equated with responsibility, and "rituals of accountability" are used not only to monitor actors' responsibility but also to construct the appearance of it (Power, 1997). Thus, emphases on accountability have come to replace distinct attention to responsibility.

Overall, the neoliberal trend of "responsibilisation" paradoxically leads to the erosion of responsibility for the collective good in two ways. First, responsibility is equated with and absorbed by emphases on autonomy and accountability. Displays of autonomy and accountability—rather than the achievement of a collective goal—are evidence of "responsible" action. Merry (2009) wrote, "[a] responsibilized society does not see individuals as socially situated but as autonomous actors making choices that determine their lives" (p. 403). Second, a previously more concrete assignment of responsibility to a centralised state is replaced by a diffuse assumption that "everyone"—the public, the state, schools, teachers, students, parents, communities—is responsible for providing high-quality education. When responsibility is diffusely located everywhere, it ultimately resides nowhere. Others have noted that neoliberalism played a role in making notions of responsibility "fuzzy," "hybrid" and in "deficit" (Bache et al., 2015; Benish & Mattei, 2020; Harlow & Rawlings, 2007), thus implying that the situation is even more extreme in a postneoliberal context.

While critical scholarship has provided a compelling account of these dynamics, it has often struggled to produce direct evidence of the powerful actors purportedly driving these changes. Nefarious intentions are frequently inferred from outcomes rather than systematically documented. Moreover, the consequences of these reforms are multifaceted. For example, the centring of the individual that elevates market-based approaches to education also calls for expansions of access and quality of education for all; in the words of Choi (2024), human rights and human capital rationales for education are "two sides of the same coin."

Institutionalist interpretations offer an alternative lens by de-emphasising the deliberate intent of powerful actors. Instead, they highlight how formal policies emerge through long-term historical and cultural shifts as well as routinised professional practices (Olsen, 2015). From this perspective, neoliberal reforms are less the product of conspiratorial design than the outcome of historical junctures, such as the collapse of the Soviet Union, which catalysed a global embrace of free markets and liberal democracy. Neoliberal models spread not only through coercion but also through mimetic and normative processes, and they eventually became seen as the most legitimate or successful approaches (Dobbin et al., 2007).

While institutional analyses have identified many of the same systemic changes as determined by critical theorists, the former interpret these changes differently. For instance, Brunsson et al. (2022) argued—analytically rather than normatively—that increased organisation, such as the proliferation of accountability structures, often results in diminished responsibility. Many contemporary accountability reforms have adopted this view and made it difficult to identify specific actors who bear responsibility. Instead, reforms tend to account for activities in the aggregate without linking them to particular authorities. For example, the earlier-cited accountability reform in Kenya (Box 2.1) sought to expand information dissemination to promote "accountability to the public" and to strengthen "social accountability" through community involvement. Similarly, a 2013 reform in Kazakhstan introduced an electronic platform aimed at improving transparency by making education data publicly available and integrating disparate information systems. While these reforms emphasised accountability, they provided little clarity regarding who is ultimately responsible for educational outcomes.

These explanatory frameworks are not mutually exclusive. It is likely that multiple technical, political, historical and institutional forces simultaneously shape contemporary trends in education policy. Future research could usefully explore the conditions under which various influences become more or less dominant. As populism and authoritarianism gain ground globally, technical and institutional mechanisms may wane.

5. CONCLUSION

This chapter provides a broad overview of evolving trends in the global script of education reform. Since the mid-2000s, the tenor of reform discourse has shifted from "autonomy-with-accountability" towards

"accountability-without-responsibility." These findings extend the analysis of Bromley et al. (2024), which, using topic modelling on an earlier version of WERD, identified a decline in general management and organisational reforms and a concurrent rise in reforms centred on data, measurement and information. The growing emphasis on data mirrors broader shifts in public administration literature towards "digital era governance" (Dunleavy et al, 2006), "new public analytics" (Yeung, 2023) and a new "regime of data analytics" (Eubanks, 2017). These patterns align with arguments suggesting that we have entered a "late neoliberal" (McGimpsey, 2017) or "postliberal" era (Furuta et al., 2023).

The trends documented here offer a foundation for future research on the changing nature of authority and governance in education. Systematically examining how patterns of accountability, autonomy and responsibility reforms vary across countries and over time would be a fruitful first step in this research. The following two chapters in this volume exemplify nuanced approaches: Chapter 3 (Verger) analyses national variations in performance-based assessment policies, while Chapter 4 (Steiner-Khamsi) highlights how global trends experience variable national resonance. Additional detailed case studies could further illuminate how national contexts shape evolving global patterns. Similarly, qualitative research is needed to understand how the discourses and practices of responsibility, autonomy and accountability are transforming. Investigations into which actors are deemed responsible, who conducts accountability practices and how data-driven reforms are reconstituting students, teachers and schools would be particularly valuable.

The transition to the accountability-without-responsibility era may undermine the status of the actor as a core locus of action. Market systems rely heavily on data and information despite their historical purpose being to support human decisions. In contrast, techno-optimists imagine new uses of data and information—tools that use artificial intelligence—that can potentially make decisions more fairly and efficiently than humans (Hagendorff et al., 2023). Data and information are increasingly valued not only for enhancing human decision-making but also for potentially surpassing human judgements in terms of fairness and efficiency by reducing subjective biases (e.g., Brayne, 2017). For instance, studies have shown that large language models can now outperform humans in basic reasoning and decision tasks, as they are less influenced by cognitive biases (Hagendorff et al., 2023). These systems offer unprecedented surveillance capabilities and can be used to monitor

vast numbers of individuals on diverse topics, linking institutional data and enabling predictive decisions that often replace human judgement (Brayne, 2017). Dunleavy et al. (2006) described how the new "digital governance" involves the integration of information and systems (rather than the disaggregation or decentralisation emphases of new public management). Data and measurement can support widespread control as much as they can assist markets and decision-making (Kabanov & Karyagin, 2018).

Finally, how responsibility, autonomy and accountability ought to be balanced in education systems remains unclear. For some scholars and practitioners, emphases on accountability threaten ideal systems of responsibility for public services in a democracy (Bovens et al., 2014; Harlow & Rawlings, 2007). Accountability overload is commonly understood to reduce performance and erode public trust in democratic governments (Borowiak, 2011; Bovens et al., 2014; Dubnick, 2011; Flinders, 2011). Actors may feel pressured to prioritise measurable, short-term results tied to accountability and then fall short on more substantial, long-term objectives and responsibilities or on those that are more difficult to measure. The challenges of allocating responsibility under datafied conditions may be exacerbated by features of the field of data science itself: decisions about creating data and information systems are (usually unintentionally) made behind closed doors and by informatics experts, sometimes without the participation of those with substantive knowledge to necessarily understand the consequences of routine decisions. In describing data science, Leonelli (2016) claimed the following:

> The distributed and global nature of data science creates challenges for evaluating the quality, import and potential impact of the data and knowledge claims being produced. This has significant consequences for the management and oversight of responsibilities and accountabilities in data science. In particular, it makes it difficult to determine who is responsible for what output and how such responsibilities relate to each other; what 'participation' means and which accountabilities it involves. (p. 1)

In other words, when accountability systems reveal that something has gone wrong, it is difficult to allocate responsibility. Moreover, it is challenging to pinpoint weaknesses or problems in the accountability systems themselves, as decisions about their construction are distributed and often reflect layers of highly technical activities. We are thus left with an opaque accounting that is nonetheless used to determine what makes a "good" student, teacher, school or system.

APPENDIX 2A

Note: This map uses contemporary borders and does not contain every reform used in the analyses, as reforms from some earlier entities (such as those from the USSR and Czechoslovakia) are also included.

Figure 2.A1 Percentage of reforms in the analyses by political entity

3. Pathways of performance-based accountability: an instrumentation perspective on educational reforms

Antoni Verger

1. INTRODUCTION

Contemporary education reforms both reflect and reinforce the use of learning data to evaluate school performance, inform administrative decision-making and improve educational outcomes (see Chapter 1 of this volume). The concept of performance-based accountability (PBA) aptly captures a core rationale of this reform agenda, whereby governments expand school autonomy while simultaneously assessing, monitoring and holding school actors accountable for student achievement. PBA is not composed of a single clearly defined policy; instead, it comprises a mix of policy instruments, each of which is rooted in different logics of state intervention. These include: a) data-gathering instruments to measure the performance of schools and individual actors – among which large-scale assessments (LSAs) play a prominent role, b) information instruments to disseminate performance results to a range of audiences and c) incentive-based measures that steer behaviour and ensure that performance targets are met (Holbein & Ladd, 2017).

PBA policies exhibit widespread commonalities yet are configured in markedly different ways across the globe (Högberg & Lindgren, 2021, Voisin, 2020). While PBA frameworks often convey an image of global standardisation, their regulation and institutionalisation often diverge significantly when they interact with existing governance structures, professional cultures and policy legacies. Their adoption and evolution are shaped not only by institutional, cultural and political conditions but also by implementation processes and the feedback effects produced along the way. As a result, an apparent global policy trend materialises

as highly context-specific configurations, in which the stakes involved, the combination of policy instruments, accountability purposes, and the roles assigned to key actors vary considerably.

Building on this premise, the chapter analyses the pathways of PBA policies in Chile, Norway and Spain, paying particular attention to how accountability instruments have been selected and combined alongside existing institutions such as school evaluations and education inspectorates. This analysis challenges the view that performance assessment and accountability tools are adopted solely on the basis of their technical merits. Instead, this chapter shows that their instrumentation is shaped by political and institutional dynamics and—at a more micro level—by processes of problem definition and the expanding, often unanticipated, uses developed by key stakeholders. Together, these factors influence the adoption, design and evolution of governance instruments at the core of PBA systems. Ultimately, PBA pathways reflect iterative adaptations to shifting goals, systemic constraints, and stakeholder pressures and interests, which steer both policy direction and the continual reconfiguration of instruments over time.

The chapter is organised into four main sections. The first section lays out the conceptual framework that guides the analysis, emphasising the theoretical basis for analysing policy pathways from an instrumentation perspective. The second section details the methodological approach, which includes a description of the comparative case study design and data collection procedures. The chapter then delves into three case studies: Chile, Norway and Spain. Finally, the discussion section synthesises the insights from these cases and draws out the broader lessons derived from the findings.

2. ANALYSING POLICY PATHWAYS: AN INSTRUMENTATION PERSPECTIVE

The concept of policy pathways captures the evolving and context-dependent nature of policies, specifically how their development is shaped by the institutional, cultural and political environments in which they emerge and unfold (Cejudo and Trein, 2023; Verger et al., 2017). Rather than advancing through pre-established, sequential phases, policies evolve through iterative processes that cut across multiple domains of public action, including public regulation and implementation. In education, these processes are further shaped by the interplay between fields

of action that are often treated separately, such as management and pedagogy, yet remain tightly coupled in practice.

Policy change, viewed from an evolutionary perspective, unfolds through both critical junctures – moments of significant disruption or decision that open windows for major reform – and incremental recalibrations of the tools, techniques and resources that guide policy action. While critical junctures can set policies on new trajectories, much of policy change occurs through gradual adaptations shaped by feedback effects, institutional constraints and shifting actor coalitions. These processes are relational and contingent on prevailing institutional logics (Bezes & Parrado, 2013; Kassim & Le Galès, 2010). Attending to these dynamics allows for a more nuanced understanding of how policies adjust and evolve in response to changing contexts and reform pressures.

To analyse these processes, this study brings together strands of public policy theory that are often applied separately. It integrates insights from policy adoption and agenda setting research, historical institutionalist accounts of policy change and continuity, and feedback theory, which emphasises the endogenous mechanisms of policy transformation. Given its focus on how instrument choice both reflects and shapes policy goals definition, the policy instrumentation approach serves as a connective thread. This approach connects institutional theories that focus on the role of organisational cultures in filtering and adapting reform pressures with constructivist approaches that highlight the influence of ideational factors and power relations in shaping policy adoption and evolution. Policy instruments are, from an instrumentation perspective, conceptualised as regulatory devices that can create new constituencies and exert pressure on existing systems, thereby shaping their enactment and ongoing transformation (Hood, 2007). Informed by this perspective, our framework considers that policy pathways are shaped by three interrelated elements: a) instruments' choice, including the definition of the problems that instruments aim to address; b) institutional and governance contexts, with a focus on their role in mediating and adapting policies; and c) feedback mechanisms that prompt policy recalibration and change over time.

2.1 Problematisation and Instrument Choice

Pathway research traces the origins of the policies and examines the factors that influence their selection and retention. It situates these processes within broader governance strategies, political dynamics and

institutional frameworks. The adoption and evolution of policy instruments are rarely driven solely by pragmatic considerations or technical rationality; rather, they reflect the influence of institutional arrangements, political and economic interests, and prevailing ideas (Maroy & Pons, 2021). As these forces interact over time, policies evolve through shifting equilibria among them, reflecting their inherently contingent and context-dependent nature (Peters, 2002; Capano & Lippi, 2017).

Viewed from an instrumentation perspective, the social construction of the primary problems and objectives that instruments aim to address is a fundamental aspect of any policy process (Cairney, 2019). This perspective requires examining not only the explicit problems identified through the selection of instruments but also the underlying representations, values and problematisations that shape policy choices. Such representations are critical in identifying which policy solutions are deemed feasible and which options are excluded from consideration (Bacchi, 2009). They also influence how policies are reformulated and evolve over time and how these decisions embed particular values, assumptions and priorities into the broader governance approach (Lascoumes & Le Galès, 2007).

Political factors ranging from partisan dynamics to strategic alliances strongly shape the promotion of specific policy instruments. These factors influence the rationale for adopting particular tools over others and help define the political projects and visions they are meant to advance. Change advocates, such as policy entrepreneurs, also mobilise specific solutions through inherently political means, often departing from purely technocratic approaches (Stone, 2012). These and other policy actors often reinterpret problems, such as a learning crisis, in ways that reflect their ideological, professional or economic preferences, and related views on how education should be governed (Rosser et al., 2022).

In sum, the choice of policy instruments, when explored from an instrumentation perspective, is a significant and complex process that intertwines technical, political and symbolic dimensions. Policy actors with the capacity to navigate the intersections of power, expertise and advocacy play a central role in the process of shaping policy decisions and influencing policy trajectories (Béland & Cox, 2016).

2.2 Institutions and Governance Frameworks

The adoption and retention of policy instruments depend on their administrative, political and economic feasibility. Instruments are more likely

to be enacted and sustained when they fit within existing administrative traditions and capacities. Research has consistently shown that administrative regimes mediate the reception of global policy trends and external pressures, and influence which policies are retained and how they are adapted locally (Peters & Pierre, 2016). The institutional foundations of public organisations influence the implementation and resulting trajectories of policy reforms. As Christensen (2006) argues, ambitious public sector reforms are often "historically inefficient" because key policy actors tend to prioritise inherited informal norms and values over the pursuit of purely instrumental goals—a tendency that becomes particularly evident when reform intentions diverge sharply from established institutional features.

National education administrations present a series of regulatory and organisational particularities that shape policy enactment processes (Bezes & Parrado, 2013). For Maroy and Pons (2021), path dependence, a mechanism through which reform efforts are constrained by past decisions, existing institutions and legislation, is key to understanding policy change and continuity patterns. Prior policy choices establish a pathway, either limiting or expanding future options. For instance, in the case of educational reforms directly affecting teachers' work, the regulation of the teaching profession, which varies markedly across the world, plays a crucial role in shaping its implementation and, more broadly, the meaning and significance of quality assurance in education (Voisin & Dumay, 2020).

Given the path-dependent nature of educational change and the deep institutional legacies shaping education systems, educational reform typically progresses through layering and sedimentation rather than abrupt ruptures. Major transformations may stem from external shocks or crises, but more commonly, old and new policy instruments coexist and interact, producing novel policy frameworks. Even when reforms fall short on intended goals, existing arrangements are rarely dismantled and rebuilt from scratch; instead, they are continuously renegotiated and reinterpreted over time (Mahoney & Thelen, 2010).

When examined through an instrumentation lens, new policy mixes can generate diverse and often unanticipated policy dynamics and outcomes, particularly when they bring together instruments that were not originally designed to work in combination. Policy mixing is closely linked to the mechanism of bricolage, which views policymaking not as a purely innovative endeavour but as an assemblage of heterogeneous elements that are not necessarily designed to fit together (Campbell &

Pedersen, 2001). For example, in some countries, teacher evaluation systems combine teacher-led formative assessments with summative evaluations based on external standardised tests. Such hybrid arrangements combine an instrumental logic of efficiency with a professional logic that enjoys greater social acceptance among teachers. While such hybridisation may help accommodate the preferences of key stakeholders and public expectations, it can also produce ambiguity and implementation challenges on the ground.

2.3 Feedback Loops: The Role of Instrument Constituencies

Policy pathways are shaped not only by path dependence or external shocks but also by policy feedback – the process by which policies themselves influence political dynamics over time. Rather than being mere outcomes of political decisions, policies act as active agents that reshape state capacities and alter the administrative landscape for future initiatives (Béland et al., 2022). When seen through this lens, policies lay the groundwork for their own implementation and evolution, influencing the likelihood and direction of subsequent reforms.

Policy feedback mechanisms can enhance state capacities and give rise to bureaucratic constituencies that advocate for the continuation and expansion of existing policies. Once an instrument is selected, it privileges certain actors and interests, thereby fostering constituencies committed to sustaining and promoting it (Simons & Voß, 2018). These groups, which are driven by political, professional and/or economic motivations, often ensure that policies remain entrenched, even if their effectiveness comes into question. Thus, policy instruments do more than serve immediate objectives; they also create enduring networks of influence that shape future policymaking.

Feedback theory, particularly its more constructivist versions, highlights the ideational consequences of policies and their role in shaping public perceptions of political legitimacy. Policies influence how individuals interpret social and political realities, including notions of equity, accountability and deservingness. As with material or resource-based feedback effects, symbolic effects can be either reinforcing (i.e., strengthening support for existing policies and contributing to their institutionalisation) or undermining, particularly when policies generate unintended or adverse outcomes that erode legitimacy or provoke resistance.

In a nutshell, from an instrumentation perspective, policy development is an iterative process in which feedback loops continuously reshape the relationship between policy design, implementation and outcomes. By incorporating both the resource-based and symbolic dimensions of feedback, we can acquire a more comprehensive understanding of policy evolution and change.

3. METHODOLOGICAL APPROACH

This chapter employs a comparative case study design informed by the methodological framework outlined in the REFORMED project's policy and politics strand (Fontdevila, 2019). The case studies reconstruct and analyse the adoption of school autonomy with accountability arrangements in distinct national contexts that rely on shared data collection and analysis tools. Data for the case studies were gathered primarily through semi-structured interviews with stakeholders, such as policymakers, civil servants, union representatives and experts. A purposive sampling strategy identified individuals with relevant expertise or roles in policy design and implementation. This was complemented by a snowball sampling method to refine the participant pool.[1] Secondary data sources, including policy documents, legislative texts and media reports, were used to contextualise and triangulate the interview findings.

This study adopts a contrasting-case design to examine how structurally distinct education systems respond to similar global reform pressures around PBA (Seawright & Gerring, 2008). As global reform discourses converge around the desirability of performance-based accountability approaches, this method enables a comparative analysis of how institutional and contextual factors shape the selection, design and evolution of related policy instruments. Chile, Norway and Spain were selected as cases because they all engage with global PBA trends yet differ markedly in their institutional configurations, political traditions and policy legacies. These systemic contrasts are evident in key areas, such as school choice, assessment regimes, school leadership

[1] While these interviews are not individually cited due to space constraints, their insights have been integrated with secondary data and academic literature to inform the analysis presented in this chapter. In the context of the REFORMED project, more than 100 interviews were conducted – 40 in Norway, 34 in Chile and 42 in Spain.

and teacher policies – dimensions that influence professional autonomy, school governance and educational practices.

The empirical analysis conducted for this chapter draws on case studies that reconstruct the policy pathways through which PBA has been adopted and adapted in each country. The analysis focuses on how governance instruments interact with existing institutions and political dynamics. Structured yet flexible coding frameworks were used to organise the qualitative data and ensure alignment with the study's theoretical lens. This evolutionary approach captures how contextual factors, policy tools and feedback effects coalesce over time to shape distinct paths to reform.

4. FINDINGS: UNRAVELLING PBA PATHWAYS IN THREE EDUCATIONAL SYSTEMS

The findings are organised around three distinct pathways towards PBA, each of which represents a unique configuration of events and policy decisions in the educational contexts analysed. For each pathway, we focus on common elements: logic(s) of instrument adoption, policy operationalisation, actors' reception (including the emergence of instrument constituencies and resistance coalitions), feedback loops and policy recalibrations.

4.1 Overlapping and Expanding Mandates: Accountability for School Improvement and Choice in Chile

Origins: from a market rationale to new purposes over time

Chile is a compelling example of how accountability in education has evolved within a highly marketised school system. Testing instruments were introduced in the country in 1982 as part of the structural educational reforms implemented by the Junta Militar government under General Augusto Pinochet (1973–1990). These reforms emphasised school choice and competition through a voucher-based funding formula and the decentralisation of educational governance. LSAs were introduced in this context following a market rationale. The initial census assessment, known as the School Performance Evaluation Test (Prueba de Evaluación del Rendimiento Escolar, PER), renamed System for Measuring the Quality of Education (Sistema de Medición de la Calidad de la Educación, SIMCE) shortly after, was administered in the fourth and eighth grades of primary education and focused on maths and

language at the time. Its primary aim was to inform families' school choices and foster competition among schools (Bellei & Vanni, 2015). Nonetheless, while the publication of school results was envisioned as a central market mechanism through the reform's neoliberal approach, it was not until the restoration of democracy in 1995 that these results became consistently public. In fact, it was a social democratic government that enacted this transparency policy, which highlights the bipartisan support for PBA in the country.

Initially disseminated through media outlets and later via the Chilean Ministry of Education's website, the publication of school scores marked a significant shift in the public role of SIMCE. Until then, assessment data focused on targeting schools that needed support and accompanying the implementation of curricular reforms. However, with the publication of school results, the assessment became central to the educational debate, and the scope and applications of standardised tests expanded significantly, incrementally integrating new functions. In 1996, the SIMCE was integrated into a merit-based payment system that attached collective salary bonuses to school performance (Mizala & Schneider, 2020). By the early 2000s, the government had adopted new compensatory programmes to encourage low-performing schools using learning data for improvement. The 2008 Preferential School Subsidy Law consolidated this policy, offering schools an additional subsidy for every enrolled "vulnerable" student (Honey & Carrasco, 2023). Schools had to accept additional accountability measures as a condition for receiving this subsidy. Schools were classified according to their SIMCE performance, and in the case of persistent underperformance, the state could impose sanctions on schools or even potentially close them (Valenzuela & Montecinos, 2017). In the context of a new quality assurance system, this accountability approach would scale up to the system level.

Policy calibrations: assessment framework responses to continuous educational crises

The fact that pro-equity school targeting policies had not addressed persistent educational inequalities, coupled with Chile's unsatisfactory performance in international large-scale assessments (ILSAs) and in SIMCE itself, prompted Chile's Ministry of Education to establish a SIMCE commission in 2003 to review the assessment policy (Parcerisa & Falabella, 2017). One of the main questions this commission aimed to address was "How can the measurement system effectively contribute to improving the quality of learning?" (Meckes, 2003). Several

recommendations from this commission were implemented between 2005 and 2009, such as a) a more detailed reporting system (e.g., presenting results based on standards and describing performance across different achievement levels); b) an increased frequency of assessments; and c) enhancing the reporting formats to tailor them to different users and training school actors in data interpretation and use (Equipo de Tarea para la Revisión del SIMCE, 2015).

The 2000s were a turbulent decade in Chilean educational politics. In 2006, a vibrant student-led cycle of protest emerged against the heritage of the education policies of the dictatorship and their effects: school segregation and marketisation, profit-making in education, the devaluation of public education and the segmentation of educational quality (Bellei & Cabalin, 2013). The government responded to this new crisis by creating a multi-stakeholder advisory board in which experts, political groups and teacher and student representatives participated. This board came up with several solutions, including the new Quality Assurance in Education system, which would be established by law in 2011. This new system reinforced PBA policies as corrective mechanisms of the market and as ways to promote more effective public oversight. However, the bipartisan nature of the committee and the fact that the political balance was fragile required that final recommendations include proposals from the political right, such as reinforcing SIMCE as a measure of school choice. The right-wing forces conceived PBA as a useful mechanism for safeguarding the school market and reinforcing freedom of choice for families. Thus, the left promoted accountability measures to counterbalance market failures, whereas the right promoted accountability to favour market dynamics. The malleability of the accountability agenda made it possible for these contradictory objectives to be met in the common course of political action. Teacher and student representatives left the advisory board and disassociated themselves from the final report's contents (Vera, 2011).

Thus, following the reforms of the 2010s, SIMCE adopted a quality assurance rationale; however, this shift did not entail dismantling the market-oriented logic that had shaped its origins and initial development. The government aimed to consolidate an "evaluative state" approach to school governance, enhancing governmental capacity to oversee education provision and causing assessments to become high-stakes tools for schools. Since then, schools have been classified into four performance categories based on SIMCE scores, student learning progress and personal and social metrics. Underperforming schools receive pedagogical

support and external evaluation visits from the Ministry of Education for four years, with continued poor results resulting in potential closure. The education quality assurance agency has worked in parallel with the Ministry of Education to make performance data more intelligible and actionable, assisting schools in drafting improvement plans through various initiatives, online tools and training seminars. With the new system, the assessed subjects expanded to include English, physical education, and information and communication technology (ICT); the levels tested expanded as well to cover six grades (second, fourth, sixth and eighth grade as well as the second and third years of middle school). All these changes came with an expansion of assessment and bureaucratic capacity, which materialised in new agencies, such as the Superintendenci*a* and the strengthened Agencia de Calidad. Schools were expected to engage in constant improvement cycles, often hiring private education consultants funded by public resources. This practice also bolstered the educational improvement and testing industry, which emerged as a powerful constituency within the PBA system (Parcerisa et al., 2020).

The dual accountability structure that emerged from these reforms, in which schools were accountable both to the state (through performance benchmarks and regulatory oversight) and to parents (through market-driven competition), generated additional pressure on schools, which perceived their autonomy as restricted. Due to its suffocating presence in the system, SIMCE faced significant criticism. High-stakes accountability measures such as teaching to the test and curriculum narrowing placed excessive pressure on schools and educators. Many teachers felt disempowered within a system characterised by flexible work arrangements, a hallmark of Chile's market model, which undermined their sense of control and professional practice. Organised resistance revealed a growing unease with the system's unintended consequences. Activist campaigns such as Alto al SIMCE and advocacy coalitions such as Educación 2020 called for a rethinking of the system. These campaigns gained traction in advocating for a shift towards less punitive and more formative assessment approaches. Their claims advanced parallel to broader social mobilisations that pressured the government to adopt ambitious reforms for regulating the market and promoting more inclusion in education.

In 2014, the government formed a team of experts to review SIMCE again. This team presented a series of recommendations that advocated for more balanced and comprehensive learning assessment, incorporating aspects such as students' personal and social development,

and promoting a more formative use of assessments. The government adopted proposals that once again expanded the assessment system by adding personal and social development tests to the assessment framework indicators, such as school climate, participation and citizenship formation, self-esteem and school motivation, and healthy life habits, thus broadening the vision of educational quality beyond academic results. New guidelines for teachers were also introduced to promote the use of learning data in the classroom for pedagogic purposes. However, not all recommendations were equally developed. The implementation of recommendations such as reducing the frequency of census evaluations and reviewing the consequences associated with SIMCE results was postponed.

The COVID-19 pandemic interrupted the administration of SIMCE, providing an opportunity for reflection on and reevaluation of the system's priorities. Even the possibility of dismantling the system was considered; however, this option was short-lived. A reduction in the number of annual tests from 11 to 7 and new improvements in the devolution system for schools were achieved; however, the return to business as usual signalled the enduring influence of entrenched accountability logics within Chile's educational governance framework (Flórez, 2023).

Summary

Ultimately, Chile's pathway exemplifies how accountability in education can evolve within a long-lived and structural neoliberal policy framework. The centrality acquired by SIMCE and the most important political groups' support for PBA measures, although conditioned to different goals, has contributed to the growth of both a bureaucratic apparatus and a burgeoning school improvement industry. Resistance from civil society and key educational stakeholders has facilitated a continual rethinking and recalibration of the system, although this has usually resulted in the system's expansion. The layered addition of purposes and uses of SIMCE – ranging from fostering school choice to diagnosis, promoting continuous improvement and increasing use by practice communities – reveals a system in which goals have accumulated rather than replaced one another. This dynamic has led to a complex and often contradictory PBA framework that exerts significant pressure on school actors and has acquired an unusual level of centrality in school governance.

4.2 Norway: Towards Quality Assurance with Professional Control

Origins: recentralisation of state control to address a learning crisis

Norway offers a compelling example of how accountability policies have evolved within a highly decentralised education system to tackle challenges primarily related to quality assurance and government oversight. The PBA pathway in Norway has been profoundly influenced by its educational system's social-democratic legacy, which has been characterised by extensive public provision, limited market pressure on schools, reliance on teachers' professional autonomy and a robust equity mandate. The deep decentralisation process that the system underwent in the 1980s – largely responsible for PBA adoption – was not primarily driven by economic efficiency but was instead aligned with the principle of subsidiarity in its aim to bring public services closer to citizens. Norway faced a learning crisis following the release of the first PISA report in 2001. The PISA results exposed significant disparities in student performance, prompting a major overhaul of the education system, which in turn instigated a wave of recentralisation aimed at enhancing quality standards and better coordinating policies across schools and municipalities. This shift did not emerge out of nowhere: the PISA shock accelerated changes that had already been under discussion for some time.

In 1988, an OECD expert commission assessed the Norwegian education policy and highlighted several critical issues, such as the absence of robust assessment mechanisms and the limited knowledge the government had about occurrences in schools (Langfeldt & Lauvdal, 2006). This report marked the beginning of a period of deliberation regarding the adoption of LSAs. At the time, external evaluations were conceived as constructive tools whose development required teacher involvement and were expected to facilitate curricular change (Augedal, 2006).

In 1997, the Norwegian Parliament requested that the Ministry of Education (MoE) develop a proposal for a national assessment system. In response, the MoE established a committee to design a framework that situated school evaluation within the broader context of quality assurance and improvement. The subsequent parliamentary debates were highly politicised, with the government being extra cautious to avoid making definitive decisions. Concerns about economic inequalities between municipalities and their capacities to implement the system emerged, further delaying progress. Shortly before the release of the PISA results

in October 2001, a quality committee was formed. The committee highlighted Norway as one of the only Western European countries without a national quality assessment system and without actionable data on students' performance. The committee advocated for changing this situation and stressed the need to link assessment with accountability, ensuring that evaluations had real consequences and that results were systematically addressed (Hovdenak & Stray, 2015).

The release of PISA 2000 in 2001 set off a series of events. While Norway's overall results were not poor, the country underperformed compared to other Nordic nations. The most shocking revelation, however, was the high level of inequality among the different student groups. For the Minister of Education at the time, the PISA results were "almost like coming home from a Winter Olympics without a single Norwegian medal." The Minister further added: "and this time, we cannot blame the Finns for being doped" (Camphuijsen et al., 2020, p. 630). Building on the findings of the quality committee, the conservative government launched an ambitious quality improvement initiative for primary and secondary education in 2003 called School Knows Best. Initially, the project included measures emphasising competition and freedom of choice; however, these ideas were significantly criticised and thus sidelined. Conversely, the establishment of the National Quality Assessment System (NQAS) – which integrated LSAs, PBA and common learning standards – garnered a broad consensus. The primary policy objective of the NQAS was to recentralise control and ensure quality within a highly decentralised system, aiming for more consistent practices across the country (Nusche et al., 2011). Assessments were conceived not only as tools to verify compliance with standards but also to promote a results-based management culture. It served as a mechanism for holding both schools and municipalities, as the primary school owners, accountable. Nonetheless, the first two years of LSAs were shrouded in controversy.

Policy calibrations: towards (thick) reflexive accountability

In 2004, Norway conducted national tests for the first time. The 2004 test results drew significant media attention and faced widespread criticism from multiple sectors. Some criticisms were technical; however, much of the policy's opposition came in the form of concerns that the tests and the publication of their results were used to overly control schools. Others criticised the tests for promoting a narrow view of education, potentially limiting schools' broader learning objectives. Additionally, the tests were faulted for adding to teachers' workloads and being scheduled

too late in the year for any meaningful follow-up to occur. Many upper-secondary students boycotted the tests in 2005, receiving strong support from student and teacher organisations. Members of parliament from the Socialist Left Party submitted a proposition to parliament in the spring of 2005, asking for a time-out in the implementation of the tests until 2007. The proposition also recommended using sample-based assessments instead of census-based ones and the utilisation of assessments for formative purposes instead of as accountability instruments. An evaluation of the assessments identified significant flaws, including poor test quality and a high level of missing values at the upper secondary level due to the boycott, concluding that the 2005 results should never have been published (NDET, 2011).

Following the 2005 general elections, a new left-green government introduced fresh political leadership and ideas to the MoE. One of this government's first actions was the imposition of a one-year moratorium on national assessments. It also decided that, to prevent school rankings, test results would only be published at the municipal and county levels and not at the school level. This moratorium marked the beginning of a recalibration of policy. The assessments were required to be valuable for supporting individual students, particularly those falling behind. To enhance their formative purpose and reduce summative implications, the administration of the tests for years 5 and 8 was moved to autumn. Significant changes were also made to the testing framework, such as aligning tests more closely with curricular competencies, introducing stricter reliability and validity standards, and developing common scales to simplify the interpretation of results for school actors. Additionally, several guides were created to help educators use test results effectively in the classroom. These changes emphasised that national tests were just one component of a broader comprehensive assessment system (NDET, 2011). In line with its modus operandi of making decisions based on experts' input, especially in critical situations (see Steiner Khamsi et al., 2020), the government adopted a more reflective, multi-stakeholder approach to accountability and saw testing as a formative tool to improve teaching practices.

New national tests for proficiency in mathematics, Norwegian and English were introduced in the autumn of 2007 for years 5 and 8. Since then, these tests have been conducted annually without interruption. However, this continuity has not eliminated debate, although, unlike in the early years of the tests, this debate does not focus on their technical quality (NDET, 2011). A key factor behind the growing legitimacy of

assessments in Norway lies in stakeholder involvement. The reconfiguration of the quality assurance system included extensive consultation with multiple stakeholders. The MoE sought input from a reference group comprised of representatives from major teacher unions, the Norwegian Association of School Leaders, the Norwegian Student Organisation, the National Parents' Committee for Primary and Secondary Education, the Norwegian Association of Local and Regional Authorities, and the City of Oslo, Norway's largest school owner (Allerup et al., 2009).

Since the new round of tests began in 2007, they have been received more positively by key stakeholders. Teachers in Norway, in contrast to what is usually seen in other countries, show favourable attitudes and opinions regarding assessments (Mausethagen et al., 2020). National authorities have invested considerable time and effort in communicating the tests' main objectives and ensuring transparency (NDET, 2011). Many stakeholders, including teachers, have even called for the tests to be made more meaningful for their work and to better fulfil their potential as educational tools by, for instance, providing more detailed information about students' backgrounds and competencies (Allerup et al., 2009).

The publication of assessment results remains one of the most controversial aspects of the PBA system, sparking politically charged back-and-forth dynamics. Initially, the data provided were quite basic and showed school results compared to the national average. This approach was criticised as being unfair because it overlooked schools' diverse socioeconomic contexts. Two years after its 2004 launch, the Minister of Education shut down the school portal and adopted a more restricted approach, delivering data exclusively to the educational community. However, in 2013, when the conservatives regained power, the publication of school results was reinstated. In 2021, coinciding with the election of a new Labour Party–led coalition, the Skoleporten was closed once again. This episode reflects continued tensions over how best to balance transparency with fairness in the education system.

In April 2022, the government appointed a quality assurance committee to review and refine the quality assessment system, emphasising academic and pedagogical quality development while reducing excessive paperwork and avoiding any micromanagement of schools. The resulting report advocates for a reflective approach, enabling schools, school owners, and national education authorities to make informed choices and meaningfully apply assessments in pedagogy. Key proposals include prioritising national assessments to identify students needing extra support

and reducing the frequency of mandatory assessments. The report also calls for the collection of more student input, particularly regarding physical and psychosocial school environments. The committee's recommendations were open for public consultation until February 2024, and the final report, once submitted to parliament, was converted into a green paper. However, its legal trajectory remains uncertain.

Local media and municipalities as emerging instrument constituencies

Despite the increasing emphasis on formative functions and a consensus among key stakeholders to reduce national-level accountability pressure, two key constituencies – municipalities and local media – have emerged as significant forces within the NQAS. These groups have expanded the use of assessments, thereby intensifying accountability pressures on schools and teachers (Camphuijsen & Levatino, 2022).

Since 2007, national initiatives, such as the Better Assessment Practices Project, have encouraged municipalities to meaningfully use assessment data and actively participate in the NQAS. Municipal councils are expected to produce annual evaluations of school performance and learning environments and, when necessary, to develop school improvement plans. In 2013, the Conservative government introduced an additional accountability mechanism: a quality standard that used the number of students achieving minimum performance levels as a key indicator in important educational areas. Municipalities falling short were subject to mandatory support from the national government. In 2016, value-added indicators for compulsory education were published for the first time, allowing for the evaluation of schools, municipalities or counties based on their contributions to student learning. These measures significantly increased accountability pressures and the role of incentive-based instruments within the PBA policy mix. Many municipalities have, in turn, passed these pressures onto schools through policies such as performance pay for principals, the publication of school performance scores or the establishment of performance-based contracts (Camphuijsen et al., 2020).

Local media have also played a significant role in increasing the amount of academic pressure faced by schools, albeit in a more unexpected and independent way than municipalities. Norway's strong tradition of local newspapers, which is partly caused by its dispersed population being located across a vast geographical area, has contributed to this dynamic. Local newspapers frequently report on school results

and rankings and publish articles criticising underperforming schools or praising those with strong results, leveraging transparency legislation to access and publish school data even during periods in which such practices are discouraged by the national government. These practices are reported to increase competition in education and exert performance pressure on schools (Camphuijsen & Levatino, 2022). Not coincidentally, the 2022 NQAS Committee recommended ensuring that test results were presented in ways that prevented local media and similar outlets from easily creating school rankings.

Summary

Adopting LSAs and related accountability measures in Norway follows a quality assurance rationale triggered by international benchmarking and competition trends in education. However, throughout its ups and downs, the Norwegian quality assurance system has advanced a school governance approach that, at the national level, remains mainly anchored in a social-democratic welfare tradition. It continues to prioritise equity goals and decentralised governance while striving to balance the introduction of new accountability instruments with the preservation of professional autonomy.

Despite the policy design intentions, significant challenges have emerged during the deployment of the quality assurance system, including teacher and school reputational pressures caused by the publication of school rankings, often by local actors, such as local newspapers and municipalities. These pressures disrupted a traditionally collaborative educational culture by introducing competitive elements to school assessment. The national government has recently sought to mitigate these effects by shifting PBA measures towards formative purposes and fostering collegial professional development. Iterative adjustments have led to a thick yet reflexive accountability approach with policies that are refined continuously to strengthen educational quality and equity objectives; however, this equilibrium remains fragile and may be disrupted by a new political shift in the national government as well as by shifting local politics.

4.3 PBA Meets a Neobureaucratic Logic: The Case of Catalonia within Spain

As a historically autonomous region in Spain, Catalonia has had the authority to implement its own educational policies since the 1980s. In

the 2000s, it pioneered the adoption of a New Public Management (NPM) approach to school governance within the Spanish educational context, introducing school-based management and professionalised leadership. The Catalan Education Act of 2009 marked a significant turning point, establishing school autonomy, strategic planning and evaluation as central pillars of governance. However, these reforms were hindered by budget cuts and political tensions.

The Educational Evaluation Decree was approved in 2010 – just a year after the enactment of the Catalan Education Act – and established the regulatory foundations for the Agency for Educational Evaluation and Foresight. This new agency was intended to operate with financial autonomy and a rotating governance structure independent of electoral cycles to safeguard the political neutrality of its work. However, its establishment was suspended in 2011 due to a combination of factors: insufficient financial resources amid a global financial crisis and a lack of political will from the incoming government to create an independent agency overseeing education (Bonal et al. 2023). Despite considerable discussion, this assessment agency was never created. Consequently, the Superior Council for the Evaluation of the Education System, a body established in 1993 with limited human and economic resources and directly dependent on the Department of Education, has remained responsible for evaluating the quality of the education system.

Administering basic competency tests (BCTs) and coordinating the implementation of ILSAs are among the Evaluation Council's primary functions. The BCTs are LSAs designed to measure students' attainment of basic skills in mathematics, literacy and science. Introduced in 2001, these census-based tests are conducted during the final year of primary and secondary education and provide insights into the overall performance of the education system. The primary purpose of these assessments is diagnostic for the educational administration and formative and orientational for schools and the broader educational community – although the delivery of the results during the summer break makes it difficult to generate meaningful discussions in schools. Furthermore, as secondary teachers design these tests, the assessments do not adhere to certain standards (e.g., item response theory) that would allow their results to be reliably compared over time. The presentation of the results considers achievement levels compared to schools with similar socio-economic compositions. The law prohibits the publication or dissemination of BCT results and does not permit schools to use them for promotional purposes

(Verger et al., 2024). Overall, this represents a low-stakes assessment and accountability policy design.

A low-stakes design with increasing uses and de facto effects

Despite their predominantly diagnostic nature, these tests have gained increasing prominence in various governmental interventions and initiatives over time. Educational inspection has played a central role in most of these efforts, incorporating these assessments into its top-down bureaucratic accountability framework. For instance, during the 2013–2014 academic year, "pedagogical audits" were introduced to address underperforming schools, as identified through BCTs. These audits provided additional resources to integrate schools into learning acceleration plans and offered targeted training and advisory services (Casserras, 2017). Subsequently, new test-related tools have been adopted to enhance school accountability, including the School Indicators System, the Annual School Evaluation, and teacher and school principal evaluations. These tools use BCT results as key benchmarks in varying ways and with differing implications. Many of these initiatives, such as the Annual School Evaluation and pedagogical audits, have not been sustained, while others, such as teacher evaluations for promotion, have yet to be fully implemented.[2] However, these efforts have reinforced the perception of BCTs as central governance tools and main proxies of educational improvement. These efforts have developed alongside a public discourse increasingly concerned with the quality of education in Catalonia. Since the mid-2010s, the region has consistently obtained mediocre results in ILSAs, often falling below the Spanish average. Educational and school success has become increasingly aligned with performance in international assessments through this discourse and related policy measures (Quilabert et al., 2024).

Educational inspection has become a key agent in expanding the use of BCTs in Catalonia and in modulating school performance pressure. As in the broader Spanish context, the Catalan inspection service

[2] For example, the inspection service plays a role in teacher evaluation and promotion. This evaluation can be conducted via two pathways: productivity (based on the contribution of teachers to learning outcomes) and the accumulation of experience/merits. Nearly all teachers opt for the latter, as the latter is highly bureaucratic and easier to manage and achieve for teachers.

operates with a highly bureaucratic logic deeply rooted in a legalistic framework, emphasising compliance with regulations and ensuring that school documentation is complete and submitted on time (Segura et al., 2021; Simeonova et al., 2020). Rarely does its work involve classroom visits or the provision of pedagogical support. Its dual mandate – to ensure compliance and support school improvement – often creates tensions and confusion among school actors. Teachers often perceive inspectors as surveillance agents rather than partners in school improvement – a perception reinforced by the limited time inspectors can spend in schools. With the emergence of LSAs, inspectors' school evaluation work has focused on reviewing inputs (such as school plans) and assessing outcomes (mainly through test data), often overlooking educational processes. Assessments provide inspectors with a valuable data source for monitoring schools' results under their supervision without requiring the development of additional instruments or the generation of new data through, for instance, qualitative observations. Inspectors require schools to include objectives, actions and indicators related to BCT measures in their management documents, signalling that results in the standardised test are a core component of educational quality (Inspecció d'Educació, 2022).

LSAs, despite their low-stakes design, have not been well received by school actors. Resistance from educators and family campaigns was particularly common during the era when Spain was governed by the Conservative Party, which was characterised by a more aggressive public discourse on performance-based accountability (Parcerisa et al., 2022). Teachers often reject national assessment data as a valid indicator of educational quality and rarely use test results to identify areas for improvement. Teaching-to-the-test practices are widespread and perceived as a strategy to alleviate performance pressures (Verger et al., 2024). The growing emphasis inspectors place on BCT has further deepened mistrust in test data and reinforced an instrumental view of external assessments. Additionally, the resource constraints faced by the regional quality assurance entity have frequently limited the effectiveness of evaluation measures aimed at fostering school improvement.

Summary: an underresourced policy framework lacking political direction

Catalonia illustrates the challenges of adopting, implementing and utilising LSAs for PBA in a context rooted in a legalistic administrative tradition and bureaucratic centralism. The politically sensitive nature of

assessment and accountability practices, combined with limited administrative capacity, has further hindered their institutionalisation and long-term consolidation. School assessment and accountability policies have followed an erratic yet incremental trajectory. The coexistence of NPM principles with traditional bureaucratic norms has created tensions and a high degree of discretion in policy implementation, resulting in uneven outcomes and patterns at the school level. Despite the low-stakes design of the BCTs, school actors experience significant pressure due to issues related to poor design and communication and the significant role of school inspectors as constituencies who expand the use of LSA data to align with their school-monitoring functions.

A significant disconnect exists regarding design issues between the intended diagnostic purpose of LSAs and their implementation as census-based annual tests in both primary and secondary education conducted in the final year of each stage. This approach inadvertently assigns a summative "aura" to the assessments and undermines their formative intent. Additionally, the delayed delivery of assessment data, combined with the limited time teachers have to analyse it, further reduces its perceived usefulness. Educational authorities fail to adequately communicate the policy's objectives and how to effectively leverage its results to teachers and schools. When considered in the context of an underdeveloped evaluation culture, these factors further deepen educators' mistrust of the external assessment policy.

Finally, in terms of incremental uses, external assessments have been increasingly employed to identify underperforming schools in need of intervention, evaluate teacher productivity (albeit with limited implementation), assess school principals' improvement plans and support data-driven school development. School inspectors have emerged as a core instrument constituency in a context exacerbated by a perceived learning crisis. They have progressively relied on external assessment data in their supervisory work, as it simplifies their tasks and reduces the need for them to obtain information from alternative sources. However, this growing reliance has also amplified performance pressure on schools and has contributed to a growing resistance to assessment and accountability measures among many educators.

In short, the LSA system originally designed to be low-stakes and primarily intended for diagnostic purposes has come to be perceived by school actors as a form of performance-based accountability (PBA). Over time, LSAs have become entangled in a broader policy mix of school improvement and accountability – one for which they were not

initially conceived and from which they are now difficult to disentangle without substantial restructuring.

5. DISCUSSION AND CONCLUSION

This chapter explores the contingent and dynamic nature of school accountability and assessment policies in education. Using an innovative analytical framework, it challenges deterministic and technocratic perspectives on policy adoption, formulation and evolution. Pathways towards PBA are neither linear nor uniform across the three cases analysed. Instead, the identified pathways reflect processes of political negotiation and institutional bricolage, where governance instruments are shaped by historical legacies, political dynamics and shifting problem definitions. The identified pathways reveal PBA as a mutating and often unstable institution. In adopting a policy instrumentation perspective, I argue that three interrelated factors shape these pathways. First, the framing and construction of educational problems significantly influence how PBA policies are justified, designed and ultimately implemented. PBA is notable in this regard: it can be mobilised to address different, sometimes competing, policy challenges simultaneously. Second, I highlight the critical role of interactions between PBA tools, large-scale assessment, and pre-existing institutional frameworks, including other accountability approaches (professional, bureucratic). PBA instrumentation is particularly sensitive to established administrative rules and norms, particularly those related to school monitoring and teacher regulation. Third, I examine the iterative nature of policy reform through policy feedback mechanisms. The outcomes of initial implementations often feed back into subsequent policy adjustments and shape distinct patterns of policy change over time. We then synthesise how these elements unfold within the empirical contexts examined in this study.

In Chile, the logic of PBA adoption has evolved from an exclusively market-driven rationale to absorbing new and overlapping purposes, including the public regulation of the market, quality assurance and equity. This layering of rationales over four decades has created a complex system that incrementally expands the assessment framework and its consequentiality over time, thus increasing the performance pressure on schools. PBA was formally adopted in Norway after more than a decade of deliberation, as a response to a learning crisis placed on the public agenda by poor performance in ILSAs. However, after some years of erratic implementation, teachers gained some control over central PBA

instruments through their institutional professional autonomy, at least at the national level. In Catalonia, Spain, the primary logic of adoption is tied to the modernisation of the education system and the promotion of an evaluation culture, but within a low-stakes design approach in which LSAs are mainly intended for diagnostic purposes. The pathway followed is shaped by the hybridisation of performance-based monitoring with a deeply rooted bureaucratic tradition of school supervision. Unclear goals, a misalignment between objectives and policy design, and an insufficient administrative capacity to develop a coherent framework and communicate it effectively have contributed to a decoupling between assessment structure and practice.

In all three contexts, instrument constituencies emerge and operate, although their nature varies significantly across each case. In Chile, the main instrument constituency consists of a school improvement and quality assurance industry encompassing both public and private actors. In Norway, municipalities and local media act as key constituencies that exert and amplify reputational pressure on schools and teachers. In Catalonia, school inspectors serve as the primary PBA constituency, increasingly leveraging assessment data to evaluate schools and benchmark their monitoring activities – a practice legitimised by recent school improvement mandates. Overall, the role of instrument constituencies helps explain why instruments expand their uses and presence within educational systems and the perception of instrument consequentiality by school actors. As suggested by the policy instruments literature (e.g., Simons & Voß 2018), these instruments evolve into "living objects" that become somewhat autonomous from the original policy decisions and intentions that created them due to the actors and interests actively promoting them.

This chapter demonstrates that while PBA policies are often introduced in response to educational learning crises, such as those produced by poor results in national or international large-scale assessments, their intensification over time can lead to these instruments triggering their own educational crises. Issues such as excessive teaching to the test, the erosion of teacher autonomy and the emergence of opt-out movements prompt governments to revisit their assessment and accountability frameworks. The typical governmental response to these challenges involves paradoxically expanding the assessment system, making it more comprehensive in scope to avoid curriculum narrowing, covering additional levels to discourage strategic teaching to the test, and training teachers to maximise productive uses of assessment data.

In both Chile and Norway, feedback mechanisms derived from implementation have explicitly driven recalibration efforts and encouraged the adaptation of PBA policies. In these countries, progressive governments have aimed to reduce the prominence of LSAs and their stakes and reinforce process-based evaluations. Despite discussions about minimising the intensity of these assessments as much as possible, the assessment systems in these contexts have generated influential constituencies and dependent users, making it challenging to reverse such policies. In Catalonia, stability might be achieved if the regional assessment framework were to align more closely with a recently adopted federal framework emphasising the diagnostic function of external assessments,[3] although it remains too early to determine the outcome of such efforts.

PBA emerges as a common response to significant challenges faced by educational systems across the three analysed countries. While LSAs and their connection to the idea of public accountability in education have received bipartisan support, the focus of this support varies depending on political orientation. Progressive forces generally advocate for combining test-based with process-oriented and reflective forms of evaluation, often expanding assessment and accountability frameworks to prioritise equity goals and formative uses. Conversely, conservative forces tend to emphasise an effectiveness-driven approach coupled with incentives and greater transparency, such as the publication of school results, often expanding the consequentiality of assessment frameworks. The centrality of LSAs within quality assurance and accountability systems varies over time (Fontdevila et al., 2024), yet major political groups, despite their different emphases, consistently conceive of these instruments as core school governance tools.

Indeed, one of the key reasons for the limited consistency of performance-based accountability (PBA) as a policy mix lies in the fact that, beyond the centrality of assessments, PBA's other core components – information and incentives – are politically salient and often highly contested. Questions such as how much information about school performance should be shared, with whom, and what types of incentives are legitimate or effective remain deeply disputed in both academia and policy arenas. More importantly, this study shows that even when

[3] Spain only adopted a national assessment framework at the federal level in 2023, following the enactment of the Organic Law 3/2020 of 29 December.

these components are formally absent or only weakly institutionalised, they can gain traction on the ground through the unofficial practices of key stakeholders. These bottom-up uses may unexpectedly intensify accountability pressures and significantly reshape how the PBA mix is perceived and enacted in real-world educational settings.

In conclusion, this chapter highlights the critical role of institutional fit in shaping the evolution of PBA tools, the role of diverse policy actors in framing educational problems and promoting specific instruments, and the iterative nature of policy development, in which feedback loops and instrument constituencies drive both incremental adjustments and more substantial shifts. By examining the interaction between governance instruments, institutional legacies, and the evolving construction of educational problems, this analysis offers a nuanced understanding of how assessment and accountability approaches both shape – and are shaped by – the contexts in which they are embedded. Future research could further investigate the dynamic relationship between the problems that policy instruments seek to address and how those instruments are received, interpreted and adapted across different levels of the system – from policymakers to frontline practitioners and targeted communities.

4. The layers of datafication

Gita Steiner-Khamsi

Tracing global trends in education over time offers important insights. However, understanding why and when a trend resonated globally, which aspects were selectively adopted, and how it was transformed to fit into a specific policy context is equally important. These macroscopic and mesoscopic perspectives arguably complement each other. Researchers in qualitative comparative policy analysis have explored the relationships, interactions, and power dynamics between actors, levels, and contexts. Therefore, a bifocal lens with a macroscopic view of global trends and a mesoscopic investigation of how these trends were received and translated in local institutions is not unusual in policy transfer research. This global/local nexus is the subject of frequent and fascinating empirical investigation and heated debate (Hartong & Nikolai, 2017).

Let us take a step back and consider the following question: Is the exclusive focus on the global/local relationship justified? What other contexts are important for understanding the changes in how data for accountability has been utilised over time and across space? In addition to exploring international and transnational relations, one might also examine the "trajectories of governance" (Capano et al., 2022) and analyse the reception and translation of global reforms across various ministries. In other words, instead of placing nations at the forefront, this chapter highlights the state and, with it, its evolving policy goals and instruments.

The investigation of sector-specific trajectories addresses the following research questions: How has the education system selectively adopted, interpreted, and reframed the various phases of datafication that have permeated public administration reforms over the past 50 years? How has the state utilised and promoted the use of data to regulate education? This chapter applies a bifocal lens that situates the global datafication trend within the broader context of public administration reform and then examines how this trend was translated into the logic and language

of education. The chapter demonstrates that, over time, the objects and objectives of datafication have evolved in response to broader shifts in public administration driven by ministries of finance and their counterparts at the global level: the OECD and the World Bank.

1. DATAFICATION IN ITS SEQUENTIAL CONTEXT

Datafication has occurred everywhere and across all sectors. In this chapter, we intend to address the semantic coupling of data with the education sector by tracing datafication in education in its own temporal and spatial context. Ecological approaches, which are pursued in sociological systems theory and historical institutionalism, are sensitive to context and lend themselves to contextual analysis. Given the ubiquity of datafication, a question arises: Who are the driving forces behind this global reform?

I propose that global school reforms ultimately be considered as being driven by ministries of finance and their counterparts at the global level, particularly the OECD and the World Bank. Subsequently, ministers of education transfer and translate these reforms for use in their respective education sectors. The current global trend of datafication in education must therefore be interpreted against the backdrop of broader public administration reforms. The contextualisation suggested here is threefold. First, it pays close attention to developments in public administration reforms. Then, it traces the sequence of the public administration reforms over time to show how each reform built on and at the same time distanced itself from the shortcomings of the previous reform to make a case for change. Finally, the ecological approach adopted in this chapter instigates an interest in understanding how institutions codify, translate, or make meaning of reforms transferred from another sector.

Looking beyond sectoral boundaries may be worth the effort. A good case in point is the global spread of participatory budgeting, which has been well documented in public policy transfer research (Porto de Oliveira, 2017). It started as a pilot project in Porto Alegre, Brazil, in 1989 and was disseminated globally with the financial support of development banks. Nearly 20 years later, the World Bank documented, promoted, and helped export the Brazilian reform in the new context of social accountability reforms (World Bank, 2008). This reform was then scaled up during the period of e-government reform, also known as New Public Governance (Osborne, 2010; see Steiner-Khamsi, 2025).

We concur with Mattei's assertion about e-governance reform (Mattei, 2023): data are used for accountability by both the state *and* its citizens. In democracies, the state draws on data to monitor publicly funded (public and private) institutions; conversely, citizens use data, rankings, and ratings to assess the quality of public services to which they are entitled. Social accountability is writ large in e-government reform. The reform aims to increase transparency through the instrument of datafication to meet the technological requirement of open access to data and information.

Second, sequential analysis (Abbott, 1995; Pierson, 2004, 2015) allows for the reconstruction of the causal chain of events leading to the evolution of datafication in public administration and subsequently, with a time lag, in the education sector. What has driven datafication, and what has changed in relation, or rather, as a reaction, to the previous reform? Such a "sequential theory" (Falleti, 2010, p. 31), recognising "reactive sequences" (Zürn, 2018, p. 89), or focusing on policy sequences (Capano et al., 2022) acknowledges that every new reform promises to usher in a new era, claims to be better than its predecessor, and asserts that it must address a long list of problems it supposedly inherited. In the wake of this process of disassociation from the past, the reform establishes new goals, change mechanisms, and policy instruments (see Steiner-Khamsi, 2025). Since institutions carry a historical legacy, are path-dependent, and change only incrementally (Thelen, 2004; Mahoney & Thelen, 2015), changes in such institutions may initially appear small and come across as a logical consequence of previous developments. Despite this, the promise of improvement hovers over every reform. A long-term retrospective view is crucial for detecting disjunctures or distinct departures from previous reform waves. Sector-specific pathways in how public administration reforms are selectively adopted, translated, and communicated exist. Our team of authors is, of course, attuned to the educational variant of datafication. Specifically, a sequential analysis reveals how datafication has been repurposed with each new reform. In the current reform, the OECD and the World Bank stress the importance of data for citizen engagement and social accountability (CESA). The reform seeks to enhance public accountability or demand-side accountability both in public administration and specifically in education, as explained in the World Bank's Foundational Learning Compact:[1]

[1] See https://www.worldbank.org/en/topic/citizen-engagement#1

> Growing evidence suggests that, under the right conditions, meaningful forms of citizen engagement and social accountability (CESA) can result in better governance, citizen empowerment, more positive and constructive citizen–state relations, strengthened public service delivery, and, ultimately, enhanced development effectiveness and well-being. (World Bank, 2025a)

Finally, the dual perspective of sociological systems theory on the processes of reception and translation (Steiner-Khamsi, 2021) can help connect the gap often created between macroscopic analyses (global trends over time) and mesolevel investigations into institutional rationales, timing, and agency of policy adoption and translation. However, unlike previous studies on policy borrowing, which focused on the global/local nexus to reveal the transnational dimension of domestic policymaking (e.g., Steiner-Khamsi, 2021), this chapter examines cross-sectoral transfer.

Table 4.1 provides a historical overview of the last four public administration reforms and specifies changes in regulation, type of accountability, and the state's role in such reforms. The last two columns present how these reforms have been adapted and translated into the education sector. More specifically, the table shows how the units of data analysis ("objects") and the purpose ("objectives") of datafication have changed over time.

2. THE TEMPORAL ORDER OF PUBLIC ADMINISTRATION REFORMS

The public sector has experienced four reform waves over the last few decades (Steiner-Khamsi, 2025), as shown in Table 4.1. It transitioned from the "traditional" model (Goldfinch, 2023, p. 3), the "bureaucratic-professional" model (Maroy, 2012) or the Weberian model (Bromley & Meyer, 2016, p. 11) to New Public Management (NPM). Furthermore, the traditional model was criticised for being rigid, hierarchical and fixated on bureaucratic-procedural rules. In practice, this neat organisational structure did not always work, as illustrated by Bromley and Meyer (2016, p. 11):

> [B]ureaucracies never worked according to the theory. The chain of command was broken at some point, as bureaucrats pursued interests of their own (now often called corruption) or reinterpreted rules in inventive ways. Or rules made no sense in particular local settings and were ignored or followed only in ritual… (Bromley & Meyer, 2016, p. 11).

Table 4.1 Objects and objectives of data collection across the last four reform periods, 1960s–2020s

Time period	Name of the reform	Type of regulation	Type of accountability	Role of the state	Objects of data collection	Objectives of data collection
1960s/70s	Traditional bureaucracy	Inputs	State-based accountability	Welfare state	Sector	Ensuring compliance checks by inspectors
1980s/90s	New Public Management (NPM)	Outputs	Market-based accountability	Entrepreneurial state	School	Promoting school choice for parents
2000s	Network governance	Outcomes	Standards-based accountability	Interventionist state	Class	Rewarding and punishing teachers
2010s/20s	E-government (New Public Governance)	Public opinion	Social or public accountability	Engaged state	Student	Enhancing the resilience and employability of students

Source: Adapted from the work of Steiner-Khamsi (2025).

Conservative political parties fought against the big and expensive state apparatus established during the welfare-state period. Their attack on the "state monopoly" was in sync with their deregulation and privatisation agenda. The NPM reforms of the 1980s and 1990s aimed to cut public expenditures and reduce the large state apparatus responsible for the redistributive functions of the welfare state. Governments hired consulting firms such as McKinsey, Deloitte and PwC to obtain advice on improving efficiency by dividing the administration into smaller units and affiliated agencies, outsourcing the provision of public goods and services and establishing contractual arrangements with the private sector. NPM and privatisation concurred simultaneously and mutually reinforced each other, triggering a deluge of businesses flooding the education sector worldwide. The role of the state shifted from redistributive to entrepreneurial.

NPM reforms were both controversial and consequential. Dunleavy, Margetts, Bastow and Tinkler (2005) demonstrated how NPM reforms devastated public administration in the pioneering country of NPM: New Zealand. By 1999, the country of 3.5 million people was left with over 300 separate central agencies and 49 small ministries. The sweeping privatisation of public services led to rising inequalities, and lowering quality standards became the hallmark of NPM.

A political move from conservative to centrist-left political parties in several leading NPM countries, notably in the UK and the United States, followed at the end of the 1990s and around the turn of the millennium. In Europe, for example, statesmen Tony Blair (Labour Party, UK) and Gerhard Schröder (Social Democratic Party, Germany) joined forces to advocate for a "Third Way" that neither followed the expensive model of the welfare state nor the laissez-faire style of the entrepreneurial state. Instead, they propagated a Third Way, which Schmidt (2009) characterized as "faire-avec" (p. 256).

These politicians reminded their like-minded political leaders to consider the negative budgetary implications of the welfare state promoted by social-democratic parties: "Public expenditure as a proportion of national income has more or less reached the limits of acceptability" (Blair & Schröder, 1998, p. 4). At the same time, they distanced themselves from the immediate past (the mindset of the entrepreneurial state) and instead proposed a Third Way:

> Within the public sector, bureaucracy at all levels must be reduced, performance targets and objectives formulated, the quality of public services

> rigorously monitored and bad performance rooted out. (Blair & Schröder, 1998, p. 4)

Third Way politics implemented governance by numbers to reregulate, at least minimally, the provision of public goods and services through standard setting and outcome monitoring. Such pro-business politics in fact actively propelled public–private partnerships, inviting businesses and foundations to involve themselves in the policy process. Not all countries pursued standard setting and target setting with equal vigour as the UK (Boswell, 2015); however, the steering mechanism shifted from output regulation (proposed in NPM) to outcomes regulation (proposed in Third Way politics), as shown in Table 4.1. The term "interventionist" state best captures the new role of the state as a standard setter and monitor of outcomes.

The interventionist state's orientation towards outcomes generated a significant amount of data at all levels of administration as well as in schools, workplaces and other interpersonal settings. Everyone was holding someone else accountable for something. Performance evaluation became ubiquitous and prominent to the extent that Lascoumes and Le Galès (2007; see also Le Galès, 2016) proposed that Christopher Hood's well-known toolkit of governments (Hood, 1983) – governance by legislation, persuasion, financial incentives, and contractual arrangements – be expanded to include a fifth tool: performance evaluation.

The proliferation of data collection, analysis, reporting and circulation was enabled by technology. The most recent public administration reform envisions an engaged state that shares data and considers itself accountable to the constituents. The digital economy, particularly the possibilities of the internet, machine learning and artificial intelligence (AI) has fuelled the state's new role and a new use of data and information. Social media has created information silos, thereby minimising public knowledge. Furthermore, as a prominent data collector, the state is pressured to justify its accumulation of data and pay rent materially and immaterially to the objects of data collection: the citizens. The e-government or "open government" engages the citizen because it must care about public opinion, which is readily and constantly shared on publicly accessible digital platforms. Data producers (citizens) should also be the data users and participate in the political process more than before. The state should engage with and listen to citizens despite the risk of succumbing, as we see today, to populist politics that counter the liberal principles of human rights. With the shift from the interventionist

to the engaged state (see Table 4.1, column 4), the state considers itself transparent, publicly accountable and responsive to citizens (Mattei, 2023; Krogh & Triantafillou, 2024).

The four reform waves discussed above have had vast repercussions for knowledge-based regulation. The shift from output to outcomes regulation (see Table 4.1) generated a warehousing of all kinds of data. Henceforth, providers could be public or private providers if they fulfilled the required standards and accreditation criteria. As with all innovations, whether good or bad, each new reform is explicit in how it intends to fix what supposedly went astray in the preceding reform. Pointing fingers at the shortcomings of and distancing itself from the previous reform is an important strategy for ensuring support for change. As mentioned before, the NPM reform relentlessly pointed to the rigidity, micromanagement and state monopoly pursued in the traditional bureaucracy and promoted freedom and customer choice as an alternative to the previous regime.

Finally, the current e-government reform criticises the previous reform for the unequal accumulation and use of data: the government should become transparent and make all its information, including data, publicly available so that citizens can hold the government accountable for its actions or inaction. A brainchild of e-government, CESA, was inextricably linked to data at the outset. The idea is to produce more data and the right kind of data that citizens or any organised non-state actor (e.g., businesses, churches, associations or philanthropies) can use to hold governments accountable for their policy decisions. The OECD and World Bank's visions of "good government" and e-government in the digital era are clearly defined and measurable: The OECD developed *Good Practice Principles for Service Design and Delivery in the Digital Age* (OECD, 2022), while the World Bank monitors and measures digital transformation in the public sector in 198 economies and has made its findings publicly available. Their GovTech Maturity Index consists of four components (World Bank, 2022) broken down into 46 indicators, of which six are for digital citizen engagement. The index measures aspects of public participation platforms, citizen feedback mechanisms, open data and open government portals.

3. TRANSLATION AND REPURPOSING OF DATA USE IN THE EDUCATION SECTOR

The nexus between ministry and policy is adequately studied in political science and public policy studies. For instance, Garritzmann and Siderius (2024) examined the ministerial logic, politics and power dynamics of three German ministries: the Ministry of Labour, the Ministry of Finance and the Ministry of Home Affairs. The authors analysed social policies passed in the Bundestag from 1969 to 2021 and found vast differences in the policy design and purpose depending on which ministry prepared the legislative document. Social policies developed by the Ministry of Labour tend to follow a social logic, whereas those from the Ministry of Finance were rooted in an efficiency logic. Lastly, policies from the Ministry of Home Affairs considered a law-and-order logic. Christensen and Hesstvedt (2024) have also greatly advanced the comparative study of ministerial policymaking. One of their recent analyses compared how various ministries in Norway utilised research evidence for policy decisions from 2000 to 2020 (Christensen & Hesstvedt, 2024). The Ministry of Education and Science ranks highest in using research evidence for ministerial white papers and commissioned green papers. More than any other ministry, the Ministry of Education also appoints the greatest number of academics as members of advisory commissions. Studies on ministerial policymaking consistently find vast differences in organisational culture and a considerable power imbalance between ministries. To reiterate the point made earlier, ministries of finance lead public administration reforms. They provide instructions on how line ministries must implement these reforms and closely monitor their integration into the institutional logic of each sector.

A triple transfer process is sometimes involved in this process: from the Ministry of Finance to a line ministry and from one line ministry to another. For example, total quality management, propelled by ministries of finance, was first introduced in the health sectors of numerous OECD countries in the 1980s before some of its key features, notably the qualification frameworks and quality assurance policies, were transferred to higher education and implemented a decade later.

When this process is applied to data for accountability reforms, the following questions arise: What did datafication mean, and how was it translated in educational terms during the three different phases of public administration reform, notably during the periods of the entrepreneurial

state (1980s/1990s), the interventionist state (2000s), and the engaged state (2010s/20s)?

3.1 Datafication During the Period of the Welfare State

To date, every education system has established norms regarding "input," ranging, for example, from normative requirements regarding school safety to providing schooling for students with special needs. In some countries, such as the United States, the government keeps these normative requirements at a bare minimum. Conversely, in other countries, the list of requirements is extensive and includes, for example, teacher qualification requirements, student/teacher ratios, and required learning supplies per classroom. In previous times, school inspectors would regularly visit schools, sometimes unannounced, to check whether schools had implemented state-issued norms. These norms were implemented across the entire sector, and inspectors collected data to evaluate compliance or deviance from these norms.

3.2 Datafication During the Era of New Public Management (1980s/90s)

The subsequent reform shifted focus away from equal inputs. NPM primarily focused on defining expected deliverables and outputs, while the input side was deregulated. Contract management gained importance as a governance tool due to new procurement laws that encouraged outsourcing the provision of public goods and services to the private sector.

Iceland is an excellent example of the datafication process set in motion as a result of inter-sectoral policy transfer (Steiner-Khamsi, Jóhannesdóttir, & Magnúsdóttir, 2024). In 1995, Iceland's Finance Ministry instructed five line ministries to pilot test contracting in their public institutions. The government at the time understood that similar public administration reforms were being implemented in other OECD countries, as pointed out in the introductory section:

> [t]he employees of the [Icelandic] Ministry of Finance familiarized themselves with how this type of agreement [contract management] is done, visited ministries and institutions in countries that have tried this method, and attended conferences of the OECD Public Management Committee on this topic. (Ministry of Finance, 1995)

The Ministry of Education was a line ministry that tasked the Women's School in Reykjavik to define and atomise core pedagogical activities and to elaborate standards and criteria against which the providers' performance could be evaluated. The Ministry of Finance and, ultimately, the OECD's Public Management Committee, to whom the government was reporting, were interested in factual information, data and measurable outputs that the Ministry of Finance could use to cost the various pedagogical "services" (e.g., teaching, grading, or meeting with teachers and parents). The information would then serve as a foundation for awarding contracts to private providers. The contracts would specify the scope of the work and quality standards. With this contracting instrument, the new public procurement law became a means to break the state's monopoly, as it had, until then, both funded and operated public schools without the involvement of the private sector.

Iceland, like other countries, revamped its government structure through the rigorous decentralisation of management and finance, followed by a shift towards school-based management. Other countries, such as the UK, the US and Chile, introduced choice and voucher schemes during the quasi-market reform era, thereby triggering an unstoppable wave of privatisation in education.

The data collected under NPM primarily served governance and financial reforms in the sector. The decentralisation of human resources and financial management at the subnational, municipal and district levels was significant. Calculating the cost per student, known as per capita financing, was intended to assist the new governing bodies with budget planning. The scope of work for providing education or "pedagogical services" was itemised, and the expected outputs were measured for contractual purposes. This allowed the government to outsource public services and goods to the private sector.

Additionally, it enabled finance departments in ministries of education to replace incremental budgeting (financial allocation based on last year's expenses) with per capita financing. This new financing modality greatly appealed to proponents of choice because taxpayer money would follow students, including those enrolling in private schools. It was the cornerstone of a new type of accountability: market accountability. This change mechanism ensures that schools will be incentivised to improve their quality of instruction because, in systems regulated by supply and demand (rather than by the state), poorly performing schools lose students and, due to per capita finance, also face dwindling funding. The data were collected at the school level to allow for parental choice.

A host of new professions was established as a result of New Public Management: professional school evaluators, data experts for designing school scorecards, "consultocrats" (Gunter et al., 2014) preparing popular school development plans on behalf of the schools and, last but not least, communication specialists in the business of "impression management" (Gewirtz et al., 2004) for schools that needed to embellish their image based on negative evaluations.

Country after country began to professionalise school management, evaluation and inspection. This global trend was fuelled by policy advice from the OECD, funding from development banks (such as, for example, the Asian Development Bank) and professional peer exchange in associations, such as the International Congress for School Effectiveness and Improvement (ICSEI). As with all global reforms, buy-in occurred at the national level, empowering some interest groups at the expense of others. For example, Swiss citizens had high expectations for quality improvement through professional evaluators, leading them to vote to abolish the century-old institution of lay school supervision. In retrospect, the district-level education authorities were the greatest losers of neoliberal reforms in Switzerland. Composed of elected laypersons nominated by political parties, such authorities used to oversee quality assurance. As in other countries, the external evaluation of schools was professionalised, rendering these authorities redundant. Formerly revered as a signpost of Swiss direct democracy in which citizens serve as honorary public servants, these authorities were replaced with professionals in the wake of NPM.

3.3 Datafication During the Period of the Interventionist State (2000s)

Network governance provided the state with a new role as a standard setter and subsequently as a monitor, evaluator, and accreditor of providers. Accreditation or certification policies mushroomed across the globe. As shown in Table 4.1, standards accountability became the new mechanism for triggering and sustaining change. Upon closer examination, the neoliberal reform had two waves: the first one promoted autonomy, deregulation and privatisation pursued under NPM (1980s/1990s), while the second wave promoted accountability, reregulation and public–private partnerships spearheaded in the Third Way political approach. Several OECD members criticised their predecessors at the turn of the millennium for their entrepreneurial approach, which enabled for-profit

providers to run schools at low cost but high profits, and for exacerbating inequality in the quality of schools in poor versus affluent neighbourhoods. Henceforth, all providers, regardless of whether they were private or public, had to adhere to common 21st-century curriculum standards. Moreover, they had to prove with data – or, more specifically with student test results – that they had implemented these standards effectively. The data were then used for teacher accountability purposes at the class level. Although highly unpopular and abolished in many countries after loud protests from teachers' unions (Jabbar et al., 2020; Levatino et al., 2024), teacher bonuses and other types of incentives were introduced on a grand scale to reward or punish teachers.

In OECD countries, the second phase of the neoliberal reform wave focused on curriculum reform and student learning outcomes. It set in motion a focus on learning, testing and standards-based accountability and a proliferation of standardised tests (Verger et al., 2019b). As with the previous reform, the OECD led and accelerated the worldwide dissemination and implementation of the reform. Strikingly, it achieved this in two ways: by defining the key competencies that should be taught in schools in the DeSeCo (Definition and Selection of Competencies) project (Salganik et al., 1999) and then by developing a standardised test to evaluate whether countries implemented the OECD curriculum framework. Reframed in terms of global governance, the Programme for International Student Assessment (PISA) may therefore be considered a reporting instrument for the education sector comparable to the Ministries of Finance's periodic reports to the OECD Public Management Committee regarding progress made in implementing new public management and procurement laws. Unlike TIMSS, Progress in International Reading Literacy Study (PIRLS) and other international large-scale student assessments administered by the IEA, the OECD's PISA was, by design, not a pedagogical instrument but rather a policy one. PISA was useless for teachers, as it did not assess the gap between intended, implemented and achieved curriculum outcomes. Instead, PISA compared alignment with the DeSeCo framework, later renamed "21st-century skills." Conversely, the IEA approach was teacher centric and meant to provide insights into improving teaching, as best illustrated in the 1995 and 1999 TIMSS video studies. In contrast, the designers of PISA appealed to economists and politicians because the test revealed their country's standing in terms of human capital stock in a competitive global knowledge economy.

In a short time, the PISA reporting instrument became a more effective driver of change than any other incremental reform. Many researchers observed the reverse mapping of curriculum reform induced by PISA: governments only began revamping their curricula after joining PISA, voluntarily reporting to the OECD and agreeing to compare themselves with other governments. Essentially, PISA assesses the performance of governments, not students. It measures whether governments have adopted and implemented a curriculum that evaluates twenty-first-century skills. The significant impact of PISA illustrates a so-called instrument constituency (Simons & Voß, 2018; Verger et al., 2019b) by which a policy instrument, such as reporting, creates its network of supporters, values and data infrastructures. The PISA test took governments by storm. By the second PISA round in 2006, it was clear that the instrument was here to stay and likely to expand its influence as its supporters sought to increase their reach. PISA marked the beginning of a new era of datafication in which data were collected to assess the performance of several public institutions, including educational systems, provinces and, eventually, schools.

Collecting and using data for performance evaluation became the hallmark of the interventionist state. As Le Galès (2016) convincingly pointed out, the choice of policy instrument is profoundly political. It reflects a specific role assigned to the state for inducing, steering and sustaining change. During the conservative era of NPM, contracting was promoted as a driver of change. During the centrist-leftist reform era of the interventionist state, performance-based measurement and re-regulation became prevalent.

It is important to remember that the objects of datafication were, at the time, publicly funded public and private institutions. In exchange for this funding, the private sector was expected to comply with the state's curriculum and learning standards. Collecting data on institutional performance helped normalise privatisation and boost public–private partnerships in education. The science of the state ("statistics") was enlisted to obstruct the the debate on whether public goods and services, such as education, should be pursued for private profit.

3.4 Datafication During the Period of the Engaged State (2010s/2020s)

A time lag is arguably necessary to retroactively assess the outcomes of a reform. The e-government or New Public Governance reform is

currently in full swing; nevertheless, it is possible to describe the reform goals, instruments of change, and thematic priorities as inscribed in the reform blueprints (see Steiner-Khamsi, 2025). CESA, described earlier as the herald of a new era, clearly reflects the agenda of New Public Governance; however, only time will tell what the outcomes of the current reform will be.

On the one hand, the current reform strengthens demand-side accountability through data sharing. Citizens – students, parents, teachers, principals and schools – are now considered both producers *and* users of data. By making evaluation results publicly available, stakeholders (rather than the state) are expected to hold those serving them, namely providers of publicly funded goods and services, accountable.

Thirty years after the sharp increase in data (e.g., school scorecards, student performance data, evaluation reports and databanks), both governments and intergovernmental organisations share a concern: their data are not used (Steiner-Khamsi et al., 2024). Digital platforms that make data publicly available have led to information overload.

For this reason, several international organisations launched initiatives in the 2020s to improve data uptake for policy, planning and practice (Baek & Steiner-Khamsi, 2024). Their attention was redirected to knowledge brokerage – that is, the linking of data production to data uptake. For these organisations, citizen engagement and social accountability should be based on data and research evidence, not on political preference. Therefore, the success of CESA depends on data uptake.

When seen from a policy perspective, demand-side accountability complements supply-side accountability measures and, if implemented rigorously, suspends the state's monitoring role in the long run. A state no longer needs to monitor the standards established during the interventionist state era, as it delegates the task of monitoring to the beneficiaries of publicly funded goods and services. It shares data with them so that the end users (i.e., the state's citizens) can directly monitor, intervene and demand good-quality public goods and services. CESA takes the principal–agent principle to a higher level and, in effect, further mitigates the state's role in initiating, steering and sustaining change. This may be interpreted as a further destatisation or reduction of the state's influence, as Ball (2018) observed for the previous reform period.

The role of technology in facilitating and accelerating the shift should not be underestimated. As shown by Stark and Vanden Broeck (2024), algorithmic management makes it possible to count, classify and reward/punish the very soul of the students, transcending the previous narrow

focus on numeracy and literacy. Similarly, e-governance or digital governance is, by default, "predictive governance," as masterfully described by Hartong, Decuypere and Lewis (2024).

Not a day goes by without new tests being developed for all kinds of skills, grades and school subjects. These tests are sold to parents and students who are worried about falling behind in today's skills-oriented workplace and society rather than one focused solely on knowledge. Given the abundance of information available today, the internet has diminished knowledge as a valuable resource and replaced it with skills, including soft skills.

4. DRIVERS AND ACCELERATORS OF WORLDWIDE DATAFICATION

International organisations have been especially attuned to soft governance by hard facts due to the lack of coercion and legislatively binding international agreements.

Therefore, the repurposing of datafication, as depicted in Table 4.1 and explained above, which continuously adds new layers of datafication, may also be observed for the first two global movers of governance by numbers in the twentieth century: the OECD and the World Bank. Each public administration reform, including its adoption in the education sector, implied a reframing and repurposing of datafication. More precisely, with each new reform, an additional layer of data was added without phasing out previous units and purposes of data collection. The fact that these two intergovernmental organisations provided the necessary legitimacy to adopt NPM, a reform that was first implemented in New Zealand, Australia, the UK, the US and Canada and then spread globally, has already been established. Mazzucato and Collington (2023) furthermore pointed out that management consultants acted as the "foot soldiers of international governmental organisations ... to export neoliberal reforms worldwide" (p. 55).

The importance of sequential analysis for understanding the rise of datafication can be illustrated by examining how the first two global movers of datafication reframed and repurposed their data collection over time. As mentioned previously, the OECD and the World Bank have been instrumental in spreading them globally. International organisations legitimise reform ideas and thereby help disseminate them globally (Dobbin et al., 2007); furthermore, their global governance mechanism of benchmarking generates a sense of urgency that accelerates the

adoption of reforms (Steiner-Khamsi, 2025). The velocity with which a reform spreads reflects the depoliticisation and decontextualisation processes that occur when the provenience of a reform is erased, which is elevated to the global level and spread as a (global) best practice. OECD countries – or more specifically, Anglophone OECD countries in the case of downsizing governments during NPM – were often early adopters of a reform that later went global (Lee & Strang, 2006). However, the temporal order of countries within a diffusion trajectory is not a sufficient cause for global diffusion.

Additionally, a quasi-external type of authorisation, or a stamp of approval by an international entity, is necessary to ensure that a country's reform spreads beyond its national boundaries. Without going into further detail here, multiple "solutions" purport to remedy the shortcomings of a previous reform. The fit with the international organisation's own mission ultimately determines which one is selected for transnational lending or dissemination. The OECD and the World Bank actively promoted each of the four public administration reforms when these reforms were in their infancy and adopted only by a few countries. Once the two international organisations became involved, convened governments – in the case of the World Bank, shareholders – and made them agree on a joint course of action, the reform spread like wildfire to every corner of the world. These organisations' technical expertise – in the form of checklists, best practices, manuals and training – and the peer pressure created by international indices, rankings and reporting requirements were only some of the mechanisms employed to ensure that government officials moved beyond lip service and implemented the reform in practice.

Reporting on progress towards achieving the agreed standards should be considered an impactful governance tool implemented by intergovernmental organisations. At the national level, this tool makes humans (inspectors) superfluous by replacing them with information systems that record and display data for everyone. As with the challenge of data uptake, the caveat to this is the resistance of many governments to rigorously collect and openly share data. This is a significant concern for international organisations whose authority rests on governance by numbers.

The rise of international data empires has occurred gradually. The OECD created the Indicators for Education Systems Programme in the late 1980s, ushering in a new era of international comparison using an ever-growing set of indicators. Like the OECD, the World Bank invested

in indicator development at around the same time. The organisation progressively expanded its databank from 116 indicators in April 1989 to 1,600 indicators in October 2018 (Seitzer et al., 2023).[2] Each established a quantification apparatus that produced data-based policy analyses and disseminated their recommendations as research syntheses. Upon closer examination, however, the purpose of data collection seems to have changed in line with broader public administration priorities.

The World Bank example shows how what governments report on and how data are used has changed its rationale and justification for data collection over the 20 to 30 years. Hossain (2023) examined the World Bank's lending for transferring Education Management Information Systems (EMIS) in low- and lower-middle-income countries, determining the period of EMIS diffusion to be 1998–2016. The World Bank implemented 311 EMIS projects in 99 countries during that period. Some countries hosted multiple EMIS projects because these either served at the country's national and subnational levels or because previous projects had failed for various reasons (e.g., computers lacking upgrades or trained staff leaving). The diffusion accelerated in the 2010s, peaking in 2013.

There are many reasons why line ministries of education are reluctant to rigorously implement education management and information systems. Many blame the EMIS for losing financial resources from international donors or their ministry of finance. On the one hand, governments must demonstrate via data that their targets and benchmarks have been met to receive tranches or final payments of loans or grants. On the other hand, statistical information is used by ministries of finance to calculate the education budget. Huge gaps between the data collected by ministries of education and ministries of finance can arise (based on national statistical offices), creating disputes that, predictably, the ministry of education tends to lose. Ministries of education typically overreport the number of enrolments, while ministries of finance underfinance the sector. When such a dispute arises, the World Bank sides with its primary clients: ministries of finance.

Regardless of the resistance to expanding the country's EMIS, the World Bank has forcefully funded the diffusion of data infrastructure in the form of EMIS at the country level. As Hossain (2023) pointed

[2] See https://datatopics.worldbank.org/world-development-indicators/stories/world-development-indicators-the-story.html

out, the bottom-up information-gathering feature of EMIS (i.e., schools collecting the data and passing it on to the subnational or national levels) strengthens schools and subnational levels of government. Thus, this feature aligns with the World Bank's larger reform agenda of school autonomy (school-based management) and the devolution of decision-making authority from the central to the sub-national level (see Hossain, 2022; Steiner-Khamsi, 2025).

In addition to three broader reform agendas – school-based management, decentralisation, and structural adjustment – the World Bank has modified its justification for building data infrastructure over time. The reasons were not primarily related to negative feedback from the recipient countries but instead changes at the global level, particularly shifting ideas about effective aid and good governance. Since the early years of the millennium, all World Bank–funded sector development projects have required baseline data, midterm targets and benchmarks regarding key indicators established by the World Bank. Statistical information is also crucial for financial planning and management, as the World Bank promotes per capita financing as part of the School Autonomy with Accountability global school reform (Verger et al., 2019a; see Steiner-Khamsi, 2025) to calculate the sector and school budgets.

However, the narrative surrounding the importance of EMIS evolved throughout its diffusion period (1998–2016): from collecting data as a management tool (including baselines, targets and benchmarks) to performance-based management (i.e., using standardised student tests to assess the outcomes of reforms) and demand-side accountability. The latter targeted those delivering services in the education sector, including teachers, school directors and local government officials, and aimed to make policy decisions more transparent, participatory and democratic in the era of e-government. In other words, the World Bank reinforced its commitment to EMIS by continuously adding new purposes and justifications for building a comprehensive data infrastructure.

In the most recent iteration of the policy narrative on EMIS, the World Bank highlights the importance of data for good, transparent governance. The World Bank actively promotes the use of data for a wide range of purposes and target groups, and it has incentivised decision-makers to utilise EMIS data to improve the efficiency of resource allocation and management of schools, teachers and communities in enforcing test-based accountability measures to improve learning outcomes (Abdul–Hamid, 2017). The World Bank has also promoted e-government,

transparent governance and demand-side accountability, all of which rely on, reproduce and analyse data.

The shifting priorities of how the World Bank has used data (from governance and planning to learning to social accountability) are fluid, overlapping and supplementary, despite coming across as repurposing data. The data infrastructure, both at the country and global levels, has become more prominent and influential with the introduction of every new utility. The previous utility was not replaced; it was given new purpose. Layering was part of the reform design (Capano, 2018). It would be more accurate to frame the process as layered datafication, as it involves adding more and more purposes for collecting and using data over time. The different layers of data utility reflect the changing priorities of the World Bank and the broader public administration reforms presented earlier. Acting as the primary counterparts of the World Bank and the OECD, ministries of finance adhere to the same reform priorities as these two international organisations.

Based on a firm belief in the workings of principal–agent theory, the World Bank has funded projects that strengthen school-based management since the 1990s. According to this framework, the principal (the government) oversees education and has agents (schools) to accomplish this task. Once the principal delegates the task to the agents in charge of implementation, the schools (acting as agents) become more motivated and efficient in carrying out the task. However, several countries encouraging parents to establish publicly funded schools (e.g., the US under Secretary of Education DeVos) may lead to a semantic shift in how principal–agent theory is interpreted. The principal will no longer be the state, but the parents, as demand-side accountability places parents, rather than schools, in the driver's seat. As briefly alluded to before, social accountability or demand-side accountability takes the principal–agent theory to new heights and, for better or worse, implements its purported change mechanism more rigorously and systematically.

5. CONCLUSIONS

Data and statistics have long been part and parcel of regulation. Foucault (1991) claimed that the very term "statistics" (i.e., the science of the state) is rooted in the state's science of knowledge-based regulation (p. 96). Starting in the sixteenth century, the French state began to count and categorise its citizens and non-citizens, first in its territory and then in overseas territories and colonies. Counting its people and their things became

"the art of government" (Foucault, 1991, p. 89) to justify its authority to rule over others beyond public contestation. Governments are not alone in resorting to this kind of science. In fact, data-based or knowledge-based regulation is the *sine qua non* governance tool for global actors.

While datafication, a device for detecting, effecting and administering policy change, has arguably existed for centuries, the objects and objectives of datafication have changed. The rise of the platform economy, which has made it possible to store, display and make incalculable amounts of data and information publicly available, has been equally important. The global trend towards datafication is evident in the growing number of laws and amendments to laws that specify the type of data used in the education sector and mandate how it should be collected, administered and utilised (Bromley et al., 2024).

Sequential analysis is necessary to understand the relationship between the various datafication waves. This chapter focused on the temporal context to elucidate how preceding events, decisions and structures shaped the scope of actions taken in the present (Tilly, 1989; Abbott, 1983; Pierson, 2004). The temporal dimension allows us to document *and* explain changes over time, interchangeably labelled the trajectories, pathways or institutional legacies of datafication. In comparison, this chapter only cursorily addressed the spatial context of datafication by looking at the first two movers of datafication: the OECD and the World Bank. The question that must be answered by spatial analysis is as follows: Which broader policy space does a government adhere to, and which are its reference societies (Steiner-Khamsi, 2025)?

Varied institutional legacies and supranational alliances carry different weights. Many international organisations are active in the education sector; however, the OECD and the World Bank have been indisputably the leaders of public administration reform. In recent years, they have collaborated with other global actors, notably UNICEF and philanthropies, to promote soft skills, including social-emotional learning (UNICEF, 2023). As with all coalition-building efforts, coalition partners join movements for their own reasons. The OECD promotes soft skills as part of its larger commitment to "non-cognitive skills" or "employability skills" (OECD, 2021; World Economic Forum, 2016), while the World Bank pursues the same route as part of the Foundational Learning Compact, as mentioned earlier. Lastly, UNICEF's mission is to support children and vulnerable groups.

Some institutional legacies last longer than others. Historical institutionalist theory contends that reforms from the distant past have a greater

impact on how institutions adopt and translate reforms in the present than more recent reforms. However, in our presentation of the last four public administration reforms (see Table 4.1), we treated the four reforms as equal in terms of impact. We agree with the assertion of Pollitt and Bouckaert (2011) that there has been a tendency in the research literature to overconcentrate on NPM reform. This preoccupation with NPM reflects the magnitude of the structural changes that NPM left behind. The role of the state was revamped and dwindled because of the reform. In comparison, subsequent reforms merely fine-tuned the reduced role of the state in regulating the education sector. Additional nuance is needed to capture the different weights of the varied reform waves and advance this line of research.

Wimmer (2023) coined the compelling term "layered legacies" to capture the multiple histories that affect how institutions respond to, translate, and implement change based on past developments. He proposed that:

> [w]e see history as a sequence of stages, each of which modifies the possible legacies left by the previous stage and shape the possibilities of future legacies through path dependency. From this point of view, we can think of history as a process of layering, with previous layers shaping the structure of later layers and later layers determining whether older layers continue to influence the present. (Wimmer, 2023, p. 2)

For Wimmer, not all past events carry the same weight in terms of shaping the reception and translation of current events. Some events are longer lasting and more consequential than others. Additionally, historical layers interact and affect each other by reinforcing, amplifying or neutralising previous experiences and legacies. In the sequential analyses, we have given equal weight to the four public administration reforms that fundamentally changed the state's role in inducing and steering change: from the welfare state, entrepreneurial state and interventionist state to the engaged state (covered in Section 3). However, NPM thinking, which was systematically pursued during the period of the entrepreneurial state (1980s/1990s), may still reverberate to this day. It would be wrong to assume that, with each reform, the previous legacy – which shaped the reception and translation of new reforms – is simply replaced by a new legacy.

Unsurprisingly, datafication in education also experienced a quantum leap during these two neoliberal reform waves. The education sector was no longer the object of reform; in chronological order, the new objects

became the school, the class/teacher and the student. Data were no longer collected for direct compliance checks but instead for steering at a distance. This distance may increase even further over time as new data-driven social accountability measures challenge the state's authority to properly regulate the provision of public goods and services.

5. Global trends in assessment reforms

Rie Kijima

1. INTRODUCTION

Assessments play an integral part in reform efforts because they can effect change and steer the course of policy decisions. Policymakers have become reliant on assessments to inform policy directions not only because assessments are useful in auditing the education system, but also because assessments advance political agendas and policies that serve the interests of those in power (Benveniste, 2002). Assessments are multifaceted and take on many objectives and functions. Governments have used assessment results to justify new reforms in the education sector. These reforms can be used to determine the pathway of children and youth entering or leaving the education system, altering prospects and opportunities in life. Assessments are typically administered to evaluate student proficiency, but they can also be used to assess teachers and their impact on student learning outcomes. Evaluation mechanisms can be used at the institutional level to assure quality and standards for comparison purposes.

Understanding global assessment reforms is important for various reasons. Assessment reforms carry substantial implications for students and their pathways for education. For example, education policy changes in entrance or exit examinations are used as mechanisms to sort students into different categories of learners. This act of sorting promotes inequality in terms of access to education provisions, which exacerbates social inequality (Domina et al., 2017). Additionally, assessment results have implications for learning conditions such as curricular content and teaching practices. If students do not perform at the level that is expected of them, then governments can introduce curricular reforms and fire teachers, thereby upending the education system and affecting millions

of children and youth (Loureiro & Cruz, 2020). The way assessments are used can also affect students and teachers. Performance-based educational policies have led to declines in the morale of the teaching corps (von der Embse et al., 2016). Studies have shown that policy measures that prioritise standardised learning and assessments have not resulted in better learning outcomes (Darling-Hammond, 1994).

This study reveals the myriad ways in which governments have articulated assessment reforms. The main research question is as follows: What are the trends and patterns of national education reforms related to assessment? This study aims to reveal the global trajectory and emergent themes associated with assessment reforms against the backdrop of the neoliberal era. We present our findings using three dimensions of analysis: macro (intersection of global and national reform efforts), meso (institutional that relate to systems and mechanisms) and micro (assessment policies that affect students and teachers). We analyse over 14,000 national education reforms that were collected from 170 countries and economies using the World Education Reform Database, Version 3, the most comprehensive database on education reforms. This chapter seeks to understand trends in assessment reform efforts during the 60-year period between 1960 and 2020.

This chapter argues that assessments are an indispensable and influential component of reform efforts, affecting student learning, teacher performance, education planning and the strategic directions of national education systems. The paper's findings highlight the rise in assessment reforms relative to other types of education reforms and the changes in typologies of assessment reforms over time. In recent years, several assessment reforms have declined in the areas of education management information systems (EMIS), quality assurance and accreditation. Additionally, the study has unearthed evidence pointing to countries referencing assessment reforms in other countries, signalling greater comparisons between countries through rankings and knowledge exchanges. Overall, the findings from this study reveal the types of assessments used to address and implement policy changes.

2. ASSESSMENTS AND NEOLIBERAL POLICIES IN EDUCATION

The liberal international order is arguably one of the most influential paradigms to emerge following the end of World War II and has affected global affairs (Nye Jr., 2017). Proponents of liberalism have promoted

policies such as adherence to the rule of law, the protection of individual freedoms and human rights, the promotion of capitalism and the free market, and accountability through democratic processes (Lake et al., 2021). As discussed in Chapters 1 and 4, the neoliberal paradigm has shifted how countries relate to one another, affecting the global world order and interconnectedness of countries around the world.

There is no shortage of literature on the influence of neoliberalism in education. Multilateral aid agencies, such as the World Bank, the Organisation for Economic Co-operation and Development (OECD) and UNESCO, have played an influential role in the construction of global education policy, thereby affecting national policy directions (McNeely, 1995; Mundy, 1999). In particular, the World Bank and the OECD have played critical roles in the expansion, promulgation and legitimisation of global and regional assessments designed to measure learning outcomes and human capital growth (Auld et al., 2019; Rizvi & Lingard, 2009). In 1997, the OECD launched its first cross-national assessment, the Programme for International Student Assessment (PISA). The OECD has, over time, played an increasingly influential role in steering countries and their national education reforms by administering cross-national assessments (Niemann et al., 2017; Sellar & Lingard, 2014). These international organisations have advocated for the implementation of neoliberal policies in education in alignment with the Global Education for All initiatives (Sahlberg, 2006).

The movement towards the standardisation of knowledge is one of the defining characteristics of the neoliberal era which began in the 1970s. During this period, countries began administering national assessments, and many became first-time participants in international assessments (Benavot & Köseleci, 2015; Ramirez et al., 2018; Kamens & McNeely, 2010). By 2015, over 130 countries had implemented a form of national assessment (Furuta, 2022), and more than 70 countries and economies had become regular participants in large-scale cross-national assessments (Kijima & Lipscy, 2024). The knowledge gained from participating in large-scale assessments has given policymakers the tools necessary to set standards, benchmark learning outcomes, and rank schools against a particular metric of educational quality, thereby advancing policymakers' ability to make evidence-based policy decisions.

The neoliberal era brought about a wave of test-based accountability reforms (Verger et al., 2019a). Policies such as high-stakes testing were adopted to incentivise and motivate teachers to teach better in the classroom; these policies also encouraged students to perform better

(Supovitz, 2009). In some cases, countries have begun implementing policies such as merit-based pay, which rewards teachers based on how well their students perform on assessments (Parcerisa et al., 2022; Podgursky & Springer, 2007). These incentives have yielded dire consequences for schools and teachers, increasing the incidence of teachers "teaching to the test" (Menken, 2006, p. 522) and having increased stress levels (von der Embse et al., 2016). In some cases, high-stakes testing has exacerbated socioeconomic inequality by increasing the gap between the rich and the poor students (Hursh, 2005).

The steady increase in the number of assessment-related reforms can be explained by the growing influence of neoliberal trends, which have propelled countries to enact reforms to increase accountability in the education sector. As shown in Figure 5.1, the proportion of assessment reforms relative to all education reforms during the 1960s was 13.19%. This figure increased to 21.98% in the 1990s and then to 31.76% in the 2000s.[1] During this period, the amount the World Bank has lent to fund education doubled from 1.5 billion USD in 1990 to 3.0 billion USD in

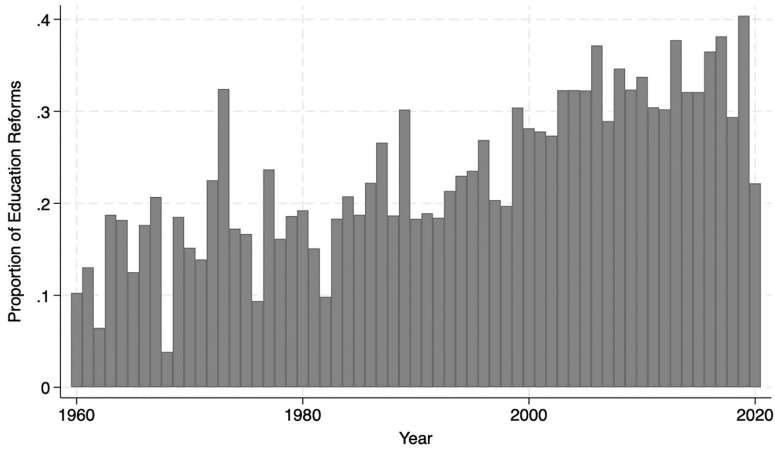

Figure 5.1 *Proportion of assessment reforms relative to all education reforms, 1960–2020*

[1] Proportion of assessment reforms relative to all education reforms by decade: 1960s: 13.19%; 1970s: 18.47%; 1980s: 20.72%; 1990s: 21.98%; 2000s: 31.76%; 2010s: 32.80% (author's calculations).

2010, shaping national education policies globally (Kijima & Lipscy, 2023). Reforms related to assessments that are used to measure equity, access and quality of education (see Chapter 2 by Bromley) are prevalent among these efforts.

3. DIMENSIONS OF ANALYSIS

We structure the dimensions of analysis using three vertical levels: global (macro), national (meso) and local/school (micro). At the macro level, assessment-related reforms are influenced by external forces and national contexts. The organisational structure that supports and enhances accountability through assessments, such as tools, mechanisms and systems that support greater accountability in education, is found at the meso level. Assessments are directly tied to reforms affecting teachers and students at the micro level.

3.1 First Level (Macro): Intersection between the Global and National

The first level situates assessments within studies of accountability in education at the macro level. This section examines the global discourses in accountability that affect national reform directions. We examine two dimensions: policy borrowing and comparisons, and national public accountability. Taken together, these two frameworks help reveal the intentions and processes associated with assessment reforms.

Policy borrowing and comparisons

We observe an increase in the legitimacy of international assessments around the world. The results of international assessments are made public internationally. These results have enabled countries to compare their outcomes with those of other countries (Martens et al., 2016). International assessments play a critical role in knowledge dissemination by referring to learning outcomes in other countries (Niemann et al., 2017). It serves as a platform for the exchange of data and analysis related to test scores. The dissemination of results from cross-national assessments has propelled countries to seek out best practices and engage in acts of externalisation (Steiner-Khamsi, 2012, 2016). By virtue of comparison, countries engage in acts of mimicking and policy borrowing on a global scale to institute changes that reflect their national political needs (Steiner-Khamsi & Waldow, 2018).

The theory of policy borrowing and comparison gives rise to three constructs that frame our analysis of assessment reforms. One of the key features of knowledge transfer is that it involves comparing and contrasting education systems against certain metrics. Governments are eager to see the ranking of countries that are published by the OECD every time the PISA results are released. Do these rankings matter? How do countries refer to these rankings when justifying their own reform efforts? Furthermore, we identify reforms that serve as a motivation or catalyst for other education reforms. Do countries reference certain reforms to justify the start of a new policy? If so, in what ways can we assess the effect of one policy on another? The third concept under this dimension is related to dependency. We seek to understand the role of multilateral agencies and non-governmental organisations in supporting, funding or advocating for reforms. These constructs are used to assess the trends in reform patterns associated with notions related to policy borrowing and comparisons.

Public accountability

Schools are part of public institutions; therefore, they are held accountable for the knowledge and skill sets passed down to children and youths. Standardised learning metrics are used to assess quality by comparing students within classrooms and between schools. Assessments keep public schools accountable for federal and state funding for education (Lindseth & Hanushek, 2009). Several controversial accountability and assessment reforms, such as value-added assessments, in which student performance is evaluated based on teacher effectiveness, have been used to determine teacher quality and link student test scores with teacher salary.[2] An example of this is the mass closure of schools directed by the Chicago Board of Education due to the schools' poor conditions and low student performance at these schools. This closure affected more than 12,000 students and was supported by the district's decision to address a financial deficit by transferring students from lower-performing schools to higher-performing ones (de la Torre et al., 2015). Assessments were used in these policy reforms to justify the shift in the allocation of public expenditure in education (Cox & Lemaitre, 1999).

[2] https://www.npr.org/2023/06/01/1178727834/after-10-years-chicago-school-closings-have-left-big-holes-and-promises-unkept (accessed on April 16, 2025).

Public accountability assessment reforms include references such as informing the public, increasing public awareness of the relevance of testing and references to efficiency in the education system. Assessments can be used to hold public officials accountable for educational quality and assessment usage in schools. In some instances, key stakeholders in education can influence the course of policy direction if the government does not meet their expectations. In Chile, students who were part of the Revolución Pingüino demonstrations referred to Chile's results in large-scale international assessments to demand better education (Kijima & Leer, 2015). When Chile skipped the Trends in International Mathematics and Science Study (TIMSS) in 2007, the government was accused of concealing the decrease in education quality. Alternatively, there have been cases of active protests in education that have led to students opting out of annual mandated examinations (Clayton et al., 2019). In 2015, approximately 25.0 per cent and 31.3 per cent of the student population in Colorado withdrew from participating in the state-wide assessment in mathematics and English language arts, respectively.[3] This trend represents heightened concerns over testing and the use of test-based accountability measures in education. Opting out of mandated assessments would skew the data and influence policy decisions, thereby undermining the credibility of assessments as a tool for accountability. Parents and students can exert their power of choice by exercising their right to refuse to participate in assessments.

3.2 Second Level (Meso): Systems and Mechanisms for the Standardisation of Knowledge

The second level of analysis reveals the systems and mechanisms that contribute to the standardisation of knowledge. The conceptual framework is derived from studies of boundary objects associated with concepts that can be translated to bridge various communities of knowledge (Fox, 2011). Boundary objects are elastic enough to adjust to local contexts and are resilient enough to carry a unique function and meaning across different contexts (Star & Griesemer, 1989). This concept is also related to the idea that localised actions are taken within a larger collective framework that justifies, rationalises and gives meaning to

[3] https://www.cde.state.co.us/assessment/cmas-elamath2015-participation (accessed April 17, 2025).

individual actions and decisions (Akkerman & Bakker, 2011). Drawing from this theory, reforms that focus on systems and mechanisms serve as boundary objects that maintain their localised adaptation while adhering to collective and normalised systems across borders. In summary, boundary objects bridge the epistemic community (e.g., universities, scientific organisations and international organisations) with the semantic communities (e.g., policymakers and bureaucrats).

The scientisation of knowledge is heterogeneous because learning is defined and assessed across different educational contexts (Star & Griesemer, 1989). The effort to standardise heterogeneity in learning is achieved through the implementation of 'regulatory governance' (Verger et al., 2019a). To put it differently, there is a need for greater standardisation to facilitate the interpretation of knowledge produced in various contexts. This study highlights two examples of boundary objects: monitoring and evaluation as well as accreditation and/or quality assurance.

Monitoring and evaluation

The trend towards scientised knowledge has justified investments in monitoring and evaluation (M&E) tools, such as EMIS and systems of evaluation (Hossain, 2023). While the term "monitoring and evaluation" is often used to follow the trajectory of a programme or a project, it frequently appears in reform efforts that aim to identify policies being articulated and implemented on the ground (Markiewicz & Patrick, 2015). This process is supported by the installation of tools and infrastructure for data collection, which facilitate the standardisation of indicators for data collection. Some examples of M&E include information systems, information technology and EMIS. Monitoring and evaluation may serve two purposes: first, to support the infrastructure necessary to keep track of data related to education, such as enrolment rates; second, to improve learning outcomes through more frequent sharing of data and analysis between schools and administrative levels (Abdul-Hamid, 2017). While EMIS gathers data that are localised and specific to each school context, the instrument facilitates the creation of commonly understood educational metrics that can then be used to inform policy decisions.

The process cannot rely solely on the development of a system; it needs to be supported by knowledge brokers who can further translate data for evidence-based policymaking (Baek & Steiner-Khamsi, 2024). This often involves global knowledge brokers who assist with the production of data, the translation of data into evidence and the use of evidence to determine policy directions to reduce the gap in the feedback

loop between researchers and policymakers (Steiner-Khamsi, Faul et al., 2024). In assessment reforms, systems and mechanisms are often linked to efforts that build the capacity to monitor and evaluate, such as enhancing their ability to administer large-scale assessments, especially for first-time participants in cross-national assessments (Kijima & Lipscy, 2020; Lockheed, 2015). Knowledge of how data are used for public policymaking is applied in various contexts through capacity building (Lubienski et al., 2014). Determining who defines quality and what kind of knowledge survives over time are questions situated beyond the scope of this study; however, the prevalence of reform discourses related to institutional capacity building suggests that knowledge brokers play an influential role in policymaking. As this study focuses on reform initiatives that revolve around monitoring and evaluation, a deeper analysis of the content of reforms is necessary to understand the type of knowledge produced or identify which actors shape policies.

Quality assurance and accreditation

Scholars have argued that quality assurance is 'not a neutral or value-free concept' (Anderson, 2006, p. 163). Researchers have claimed that quality assurance is a system that promotes ideological discourses that are impacted by stakeholders and key actors in higher education (Salter & Tapper, 2000) and is defined by the politics of governance in the production of knowledge, management of knowledge and exertion of control over what constitutes quality in education (Morley, 2003). Quality assessment is a highly politicised act; only a handful can exert their power to determine the parameters of quality. These parameters, which are set by a selected number of experts, can undermine the quality of learning provided by various educational institutions.

Quality assurance and accreditation are tools utilised to advance a certain notion of quality, especially when the number of stakeholders providing educational services increases (Altbach et al., 2010; El-Khawas, 2007). These quasipublic entities have become more prevalent actors in quality assurance and accreditation, especially in tertiary education. Today, there are more than 300 members of the International Network for Quality Assurance Agencies in Higher Education that develop accountability policies for higher education institutions.[4] The need for quality

[4] INQAAHE website: https://www.inqaahe.org/ (accessed 1 April, 2025).

control through validation of educational programmes and certification of degrees is significant considering the expansion of the higher education sector (Dill, 2010). Oversight and regulation influence the provision of education, thereby impacting the supply and demand for educational opportunities. Policymakers have sought to control and set quality standards through greater oversight and regulation while promoting the decentralisation and privatisation of educational services.

3.3 Third Level (Micro): Assessments as Measurements of Learning

The third level of analysis focuses on assessment reforms that are administered in classrooms. This level's primary objective is to measure learning outcomes and set learning standards. This section highlights the importance of assessment reforms that affect students and teachers.

Student assessments

Countries that participate in large-scale international assessments experience a drop in their test scores, which creates havoc and spurs public debate over failed policy reforms and unsuccessful educational interventions (Waldow, 2009). Student assessments, including the training and hiring of competent and effective teachers and the impact of the revision of national curricula, are used as mechanisms for increasing accountability at the government level. While the primary function of student assessments is to evaluate student proficiency, it also sets standards and metrics for comparison and competition, devolves decision-making and defines the parameters for educational quality.

This section identifies the most prevalent student assessment reforms identified in the analysis. First, assessments are used to set standards of proficiency, such as identifying the minimum passing grade, certifying students for their specialised skills and determining who receives remedial education. Assessments function as a metric by which knowledge is standardised and students and schools are compared and ranked (Guskey, 2007). Second, student assessments restrict or expand pathways for higher levels of learning. Student assessments such as entrance and exit exams are common forms of assessment that restrict access to education. Third, assessment reforms point to the rise in importance of competency-based learning or student-centred learning with an emphasis on individual pace and process towards mastery of knowledge (Henri et al., 2017). This type of assessment constitutes alternative ways of assessing

students' knowledge. The next section explores assessments that are related to another key stakeholder affecting student learning: teachers.

Teacher assessments

How effectively a teacher teaches in the classroom is an area of great concern to policymakers around the world. Reform success is highly dependent on how well teachers can adapt to new changes and demonstrate their ability to teach effectively using new curricula. Several studies have shown that teacher effectiveness can no longer be evaluated using traditional definitions of teacher qualifications, such as pre- and in-service training, credentials, degrees and years of experience (Darling-Hammond, 2010). Instead, teachers must demonstrate their teaching competence by applying their content knowledge, adapting it to different types of learners and iterating on their pedagogical approaches.

A teacher's qualifications are rarely linked to their ability to teach well (Darling-Hammond, 2010). To close this gap, countries strengthened their emphasis on instituting education reforms that focus on assessments for teachers and assessments that support teachers. The extant literature points to various modalities of assessment. In this study, we highlight the three factors most pertinent for understanding teacher reform and assessment policies. The first category is related to teacher performance assessments, which help teachers identify the gap between competency and classroom performance. Through the implementation of classroom observations and extensive feedback, teachers can apply the lessons learned to become more effective and impactful teachers (Darling-Hammond, 2010). The second category is related to assessments for teachers to audit teacher quality and their teaching practices. Studies have shown that teachers' content knowledge and performance on standardised exams are positively correlated with student learning outcomes (Santibañez, 2006). The third category consists of assessment reforms that increase teachers' assessment literacy (Koh, 2011). These types of assessment reforms are often associated with teacher training and professional development opportunities to help teachers become consumers of student-level data to inform teaching practices. The three levels of analysis are shown in Box 5.1.

BOX 5.1 LIST OF THEMATIC AREAS AND DIMENSIONS OF ANALYSIS

First Level: Public Accountability

- Policy borrowing and comparisons: a) a reform in another country as a motivation, catalyst or start of the reform; b) comparison of results with other reforms implemented in another country; c) support or funding from another country/organisation
- Public accountability: a) informing the public; b) evaluation as a form of accountability and transparency

Second Level: Systems and Mechanisms for Assessment Reforms

- Monitoring and evaluation: a) development of a new measurement, standard and metric for learning; b) infrastructure for data such as EMIS; c) institutional capacity to administer their evaluation
- Accreditation and quality assurance: the establishment of a system of accreditation and/or quality assurance at schools/universities, regional offices or federal/governmental organisations

Third Level: Assessments as a Measurement of Learning

- Intends to evaluate the performance or knowledge of students
- Intends to evaluate the performance or knowledge of teachers

4. DATA AND METHOD

This study utilises reforms from the third version of the World Education Reform Database (WERD).[5] This database contains data on 14,375 reforms from 216 countries and economies. The reforms that appear in

[5] WERD can be downloaded here: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/C0TWXM

WERD were identified from national documents produced by multilateral aid agencies, such as the World Bank, the OECD and UNESCO as well as the International Encyclopedia of Education (Bromley et al., 2023).[6] The database consists of a unique reform ID, a brief description of the reform and the reform measure. The coding of the reforms was conducted in two phases.

In the first phase, 10,236 reforms from WERD Version 2 were hand-coded by policy area, subject, level of education and thematic area. We identified education reforms that were coded as a "yes" response to the following five thematic areas: a) policy area: assessment and student learning standards – national (Q5); b) policy area: assessments and student learning outcomes – regional (Q6); c) policy area: assessment and student learning standards – international (Q7); d) monitoring and evaluation (Q10); and e) education quality (Q28). This process narrowed down the data to 4,536 education reforms.

The second phase identified 8,275 new education reforms that were added to WERD Version 3. These reforms were then combined with data from the first phase, yielding a total of 12,811 education reforms. A coding document was created to identify reforms related to assessments. The coding document consisted of six key constructs and a total of 20 questions. The following constructs were selected for their relevance to the analysis of assessments: a) assessments and quality, b) beneficiaries, c) purpose of assessments: public/political accountability, d) purpose of assessments: intentions and acts, e) policy borrowing and comparisons, and f) impact of assessments. In both phases, reforms were hand-coded according to these dimensions of analysis, and each reform could be coded for multiple thematic areas. The intercoder reliability for hand-coding in phase one and phase two was 0.8 and 0.9, respectively (Cicchetti, 1994).

The findings report the proportion of assessment reforms relative to the total number of education reforms to account for variations in reforms over time and across the number of reports from which the reforms are derived. All figures report findings using a non-parametric method – lowess – that shows a smooth curve through data points to simplify and better visualise the underlying trend (Cleveland, 1979; Fox, 2000). All reforms prior to 1960 and after 2020 were dropped from the

[6] The list of reports is provided in Chapter 2 by Bromley.

database to ensure that the analysis focused on the last 60 years of data (from 1960 to 2020).

5. KEY FINDINGS

5.1 Result 1: Overall Trends across Three Levels

The findings indicate an exponential growth in the number of assessment-related reforms between 1960 and 2020 (Figure 5.2). The proportion of systems and mechanism assessment reforms was most prevalent, followed by assessment reforms related to students and teachers. Reforms associated with public accountability were the least prevalent. Reforms associated with systems and mechanisms (level 2) increased from 10 per cent of total education reforms per year to 40 per cent in 2019. A similar level of growth was observed for reforms associated with assessments related to students and teachers (level 3). Reforms related to the measurement of learning increased from 5% in 1960 to 35% in 2019. The proportion of reforms associated with public accountability (level 1) increased from 2% in 1960 to 18% in 2019.

The findings point to the prevalence of assessment reforms that aim to develop systems or mechanisms. These reforms have enabled countries to develop the EMIS system that has never existed before. These systems provide a foundation to ensure that the government can set standards, collect data and engage in knowledge dissemination. Establishing a robust assessment system allows governments to collect more data, which improves their capacity for evidence-based decisions in policymaking. The findings point to a gradual upward trend in public accountability measures. These types of reforms have not increased as much as other types of assessment reforms.

There are some plausible explanations for this trend. Open societies exhibit a greater tendency towards these actions than countries that are characterised by a closed society. It is possible that the countries that are enacting reform assessments are emergent economies that rank lower on indicators of liberal democracy. Studies have shown that regime type matters in explaining the government's willingness to disseminate data (e.g., Hollyer et al., 2011). Further analysis using liberal democratic indices may reveal more explanations for fewer reforms related to public accountability than for other types of reforms. Additionally, several studies have shown that civil society organisations can gain support from multilateral agencies to exert greater pressure on their national

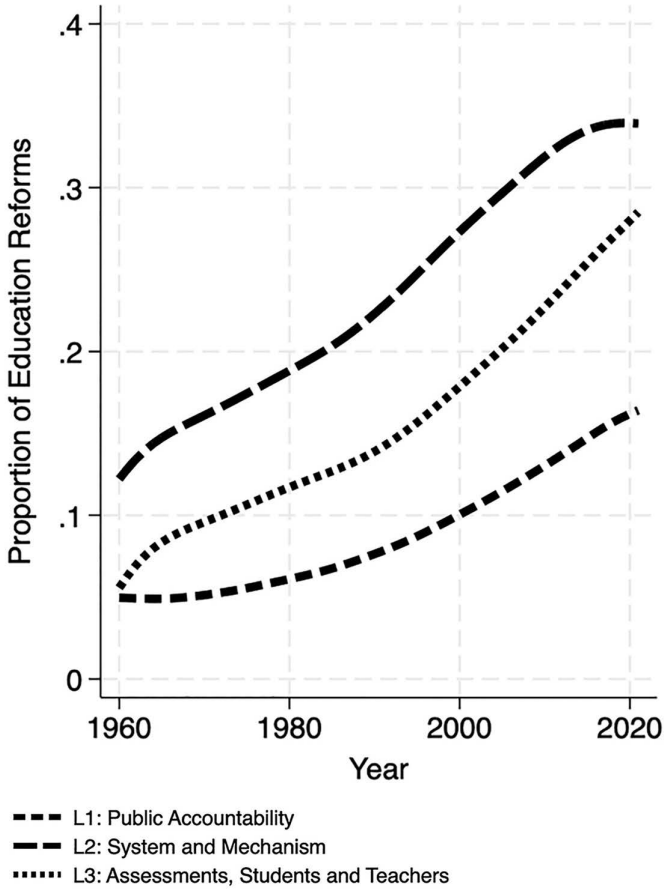

Figure 5.2 Proportion of assessment reforms by three levels

governments to enact change. An analysis of this dynamic, known as the "boomerang effect" (Keck & Sikkink, 1998, p. 19), reveals the strength of transnational advocacy networks and their promotion of greater transparency and accountability. Furthermore, the private sector has played an increasingly influential role in the regulation and provision of education in areas of accountability, especially during the neoliberal era (Abdul-Hamid et al., 2017). The sector's involvement in public–private partnerships and subsidising educational services can also explain the

shift in discourses related to accountability reforms. Our findings on assessment reforms that focus on policy borrowing and comparison are explored in the next section.

5.2 Result 2: Results of Policy Borrowing and Assessment Reforms (Level 1)

This section reports findings that focus on constructs related to policy borrowing in assessment reforms. A positive and upward trend in countries that reference other countries' education reforms associated with assessments (Figure 5.3) can be observed. Each reform measure under this dimension was coded for three concepts: compare, catalyst and dependency. The focal question for compare was whether the reform mentions or compares results with other reforms implemented in another country. For catalyst, we identified reforms that mentioned another

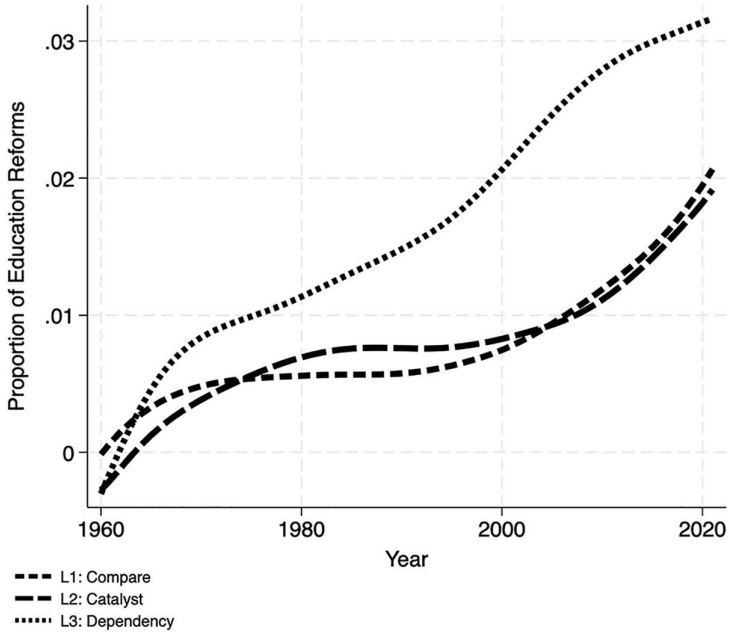

Figure 5.3 *Proportion of assessment reforms related to policy borrowing*

country's reform as a motivation, catalyst or start of a new assessment reform. In our analysis of dependency, we enumerated reforms that mentioned obtaining support, either in the form of advocacy or technical assistance or financial commitments from another country or organisation associated with the reform.

Comparison

Further analysis was conducted to identify assessment reforms that compared the results with those of other reforms implemented in another country. The findings point to three subcategories under this category: 1) reference to international standards set by multilateral aid agencies, 2) regional diffusion of knowledge and practices, and 3) justification of their reform through externalisation.

The first type of comparison reform references international standards set by multilateral organisations. For example, Albania refers to support obtained by other OECD countries and other international organisations when reforming their vocational education system and its management structures, including teacher qualifications and curricula (Albania #67).[7] Another example is from Nigeria: "To achieve the long-term broad objectives, one of the challenges is for the nation to raise the quality and standard of education to international comparative levels" (Nigeria #3715). In some instances, countries refer specifically to international assessments such as TIMSS, PISA and TALIS.

The second type of comparison reform refers to the regional diffusion of knowledge and practices. Notably, some countries, such as Ukraine, have specifically mentioned other countries in their reforms. Referring to their first-time participation in TIMSS in 2007, Ukraine was compared to other neighbouring countries such as the Russian Federation, Kazakhstan, Armenia and the Baltic countries (Ukraine #6810). Another example is when a lower-ranking country compares itself to higher-ranking countries in international assessments. Cote d'Ivoire referenced Japan, South Korea, Shanghai and Singapore when rationalising its efforts to improve its quality of education (Cote d'Ivoire #1620).

The third type of comparison reform signals externalisation. Denmark introduced a new national assessment system in 2006 "as a consequence of

[7] Education reform measures derived from the WERD database are indicated by the country name and the reform number, which corresponds to the unique reform ID number assigned to each reform.

the PISA result somewhat below the expectations and the OECD report" (Denmark #1893). Another example is a reform from Azerbaijan implemented after the PISA 2006 scores were released. The results revealed that Azerbaijan ranked among the lowest of all OECD partner countries.[8] They referenced their poor performance in PISA and their desire to deliver education of higher quality (Azerbaijan #10072). Governments implementing these reforms acknowledge that their reforms could be improved and that referring to other countries to justify the launch of their new education reform is a valid strategy.

Catalyst

Reforms that were sorted into the catalyst category refer to reforms in another country or region as a motivation, catalyst or start of the reform. This concept is directly related to borrowing ideas from other countries to influence their own reforms. The analysis revealed two typologies of catalyst reforms related to assessments: 1) reference to regional frameworks and initiatives, and 2) identification of best practices.

First, assessment reforms articulated by European countries reference European and regional initiatives. In Armenia, the Bologna process has played a catalytic role in the establishment of a national system for quality assurance at the tertiary level (Armenia #263). In Kazakhstan, the signing of the WTO agreement and the Lisbon Convention prompted the country to introduce a reform related to the quality assurance system at an international university located in Astana (Kazakhstan #673). Participating in multilateral agreements and initiatives has encouraged countries to adopt certain reforms to meet international standards.

Another type of catalyst reform refers to the successful adoption of assessment reform policies in other countries. Countries emulate practices in other countries to justify their decision to adopt similar reforms. An example of the adoption of best practices can be found in the reform articulated by the Netherlands:

> In 2012, the Netherlands followed the example of strongly performing education systems such as Australia, Ontario (Canada), New Zealand and Scotland with the development of a teacher register, which will be mandatory from 2017. (Netherlands #5039)

[8] Azerbaijan ranked 34th in mathematics, 55th in science, and 54th in research out of 57 countries that participated in PISA in 2006.

The reference to best practices can also take the form of policymakers learning from other best practices and obtaining information and policy ideas through official visits. Singapore uses "visits and study of school systems in several other countries" (Singapore #5613) to determine the kind of reforms to be implemented at the secondary level. These informal networks also contribute to exchanges of knowledge and the adoption of similar or related reforms across countries.

Dependency

The analysis also included references to reforms that mention support from other countries or organisations. This measurement is more prevalent than catalyst and comparison reforms, indicating the role of aid agencies in supporting assessment reforms around the globe, especially in the Global South. The most frequently mentioned organisations include the World Bank, one of the most influential international organisations financing education projects; the OECD and its assessments, such as PISA; and UNESCO/UNESCO IBE, which specialise in education sector planning and support. UNICEF also appeared frequently in the analysis due to its technical assistance in either implementing a new project or supporting ongoing reforms in schools and communities. Many countries also referenced the International Association for the Evaluation of Educational Achievement (IEA) and their cross-national assessments, such as TIMSS and Progress in International Reading Literacy Study (PIRLS), albeit less frequently than the OECD, their competitor in international assessments. Other regional multilateral aid agencies, such as the Asian Development Bank and the Inter-American Development Bank, also appeared several times in the analysis. Notably, regional initiatives, such as the European Commission and the Bologna Declaration, were referenced by countries that provided support to countries looking to administer and implement their assessment-related reforms.

These findings reveal a clear trend. We observe evidence of countries articulating reforms that reference achievements and standards related to assessments in other countries. While there were virtually no compare, catalyst or dependency reforms in the 1960s, these reforms have become more prevalent in recent decades, with a noticeable rise since the 1990s, which coincides with the start of the World Education for All movement. References to other countries, aid agencies and regional initiatives have been conducted intentionally and frequently since 1990. Articulating reforms enables countries to obtain support for their existing and new reforms that need political buy-in. Reforms are not automatically

transported to another country; countries articulate reforms because there exists a domestic or international condition that is conducive to the articulation of reforms (Steiner-Khamsi, 2021). The act of referencing other countries and agencies is advantageous because it allows the government to gain financial, institutional and political support, as incentives, motivations and outcomes are tied to reform announcements. These reforms are oftentimes associated with multilateral agencies, including organisations that administer international assessments, such as the OECD and IEA. This study's findings reveal the influence and power that these agencies exert over national policy discourses and reform directions that steer the path for how education is managed around the world. The specific conditions for policy adoption are an area for further analysis. These findings point to the critical role multilateral organisations play in rationalising and justifying reform articulation to signal or affect changes in national education systems, which also shape the contours, dynamics and tendencies in reform patterns around the world.

5.3 Result 3: Systems and Mechanisms (Level 2)

This section focuses on the systems and mechanisms that support assessment reforms. This is the largest category of reforms associated with assessment reforms, which consists of 1) monitoring and evaluation (M&E), 2) institutional capacity building, and 3) quality assurance and accreditation. This study reveals that the most prevalent reforms under this category are monitoring and evaluation (M&E). The second most frequent occurrence of assessment reforms involved institutional capacity building, followed by quality assurance and accreditation (Figure 5.4).

Monitoring and evaluation

M&E reforms are implemented within government structures, such as ministries of education or departments that oversee EMIS systems or data management, to improve their capacity for greater oversight of their education system. One example of such reforms is Azerbaijan's reform in 2003, which referenced the "establishment of the EMIS and Strategic Analysis, Planning and Personnel Units at the Ministry of Education" (Azerbaijan #11456). This type of institutional-level change requires additional resources and other systems for management, such as the setting up of a new data management software or the utilisation of an international data management platform to consolidate data at various

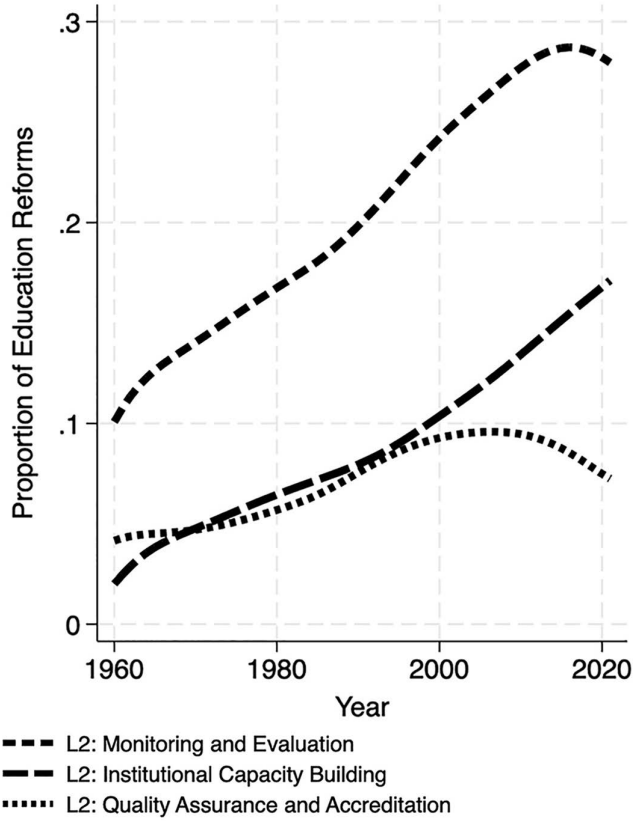

Figure 5.4 *Proportion of assessment reforms related to systems and mechanisms*

administrative levels, such as schools and provincial, regional and central governments. We observed a decline in the proportion of M&E reforms in more recent years.

Institutional capacity building

The findings point to a gradual increase in system-related assessments focusing on institutional capacity building, such as increasing the technical capacity of higher education institution leaders in data management and oversight. For example, in Yemen, a higher education law led to the

establishment of the Higher Education Quality Assurance Council. This reform also specified that it would provide "capacity building in the field of university governance and management, strategic planning, sector financing, quality assurance and quality management" (Yemen #9320). Technical support and training typically accompany large-scale reforms that involve the implementation of a new governance structure affecting the system of evaluation.

Quality assurance and accreditation

Our findings also reveal moderate and incremental reform efforts to improve quality assurance or accreditation efforts until the early 2000s and then a decrease in the last decade. Accreditation and quality assurance policies range from the establishment of an independent accreditation system to the external evaluation of higher education institutions (Belgium #668, Chile #1303, Finland #2423, and Indonesia #3125), the standardisation of university entrance examinations (Armenia #246) and the accreditation of academic institutions (Egypt #2097). Quality assurance and accreditation are often conducted by external experts and personnel with varying standards for what could be considered acceptable quality. Additionally, this process is associated with the establishment of a new accreditation agency or organisation that is affiliated or unaffiliated with government agencies. The findings indicate a decline in assessment reforms related to monitoring and evaluation and quality assurance and accreditation mechanisms.

The reasons for the decline in M&E as well as quality assurance and accreditation can be explained by two possible factors. First, the installation of an M&E system may not necessarily result in better practices that improve the context of learning in the classroom. EMIS strengthens the capacity to collect data at various levels of governance – in schools and between schools and at the district, regional and central levels. While data are captured and organised more efficiently through the EMIS system, the establishment of EMIS is not directly associated with learning outcomes, as determined by test scores. Many other intermediary factors affect learning outcomes, such as classroom instruction, total hours of instructional time, teacher quality, curricular content and student characteristics. M&E does not equal quality improvements. Second, there are reasons to believe that the establishment of EMIS does not improve efficiency in the education sector. The World Bank highlights various challenges and issues associated with the usage of EMIS, including data management, financial sustainability, and the role of the private sector

(Abdul-Hamid et al., 2017). Financial and human resources are necessary to establish and maintain a robust, effective, functional and efficient monitoring and evaluation system. Approximately 1–7 million US dollars are needed to set up the EMIS and to maintain it over the course of a 3–4-year project cycle.[9] Without the support of external funding and adequate support for maintenance and upgrades, these national EMIS systems will quickly become obsolete and unsustainable. These factors may have contributed to the decline in reforms that are specifically related to the development of EMIS.

As theorised in the introduction chapter, the overall decline in assessment reforms related to quality assurance and accreditation can also be attributed to the rise in the complexity of various stakeholders and actors involved in quality assurance (Lubienski, 2019). The rapidly changing landscape of actors in the provision of educational services, including both the private and public sectors, necessitated the government to establish regulatory frameworks to control the quality of education, especially at the tertiary level. This is related to efforts to devolve and decentralise power and governance while providing oversight and control over degree-granting academic institutions. The involvement of the private sector has also elevated threats to the education system, including fraudulent activities and non-adherence to certain standards through bribery and corruption (Sanyal & Martin, 2007). These concerns have increased the demand for surveillance and a new structure for governance to meet the rapidly changing and expanding education sector.

Governments have responded to the rising complexity of quality assurance in various ways. First, governments have instituted more regulations and frameworks for greater oversight and control. This strategy requires additional reforms, resources and systems to control and manage the burgeoning education sector. Singapore is a great example of this – the country aims to retain control over quality standards through strong oversight of its accreditation and quality assurance systems (Lo, 2014). Many countries across Latin America and South Asia have seen a rapid expansion of privatisation in education, with non-state actors in

[9] Global Partnership for Education (2022). What can we learn from EMIS Diagnostics? https://www.globalpartnership.org/blog/what-can-we-learn-emis-diagnostics:~:text=And%20these%20investments%20often%20represent,US%247%20million%20per%20project (accessed on April 1, 2025).

education providing – either fully or partially – accreditation systems and quality assurance mechanisms (UNESCO, 2022). For example, in Bangladesh and India, a designated national public university oversees the accreditation of newly established private universities and ensures that regulations are followed in subsequent years (UNESCO, 2022). On April 24, 2025, the US government signed an executive order overhauling the existing accreditation system in the higher education sector, fundamentally altering the way universities obtain more than $100 billion in federal funding.[10] This is a sweeping change from the earlier system, whereby college accreditors wielded greater independence and authority over the process of evaluation and deliberations associated with the accreditation process.

5.4 Result 4: Comparison of Student Assessments and Teacher Assessments (Level 3)

The final set of analyses focuses on assessment reforms related to students and teachers. The analysis of student and teacher assessments revealed an overall increase in the amount of assessment reforms related to the quality of learning in classrooms (Figure 5.5.).

Assessments: Students

Students are assessed on a regular basis to identify what they understand and what they do not. The analysis includes reforms related to three functions and purposes of student assessments: 1) improving the standard of learning, 2) altering the pathways for learning, and 3) evaluating students and their competency-based learning.

Standard of learning

The post-EFA era (1990–2020) is characterised by standard setting and standardisation of knowledge. This period saw many references related to minimum standards of knowledge. In Indonesia, a new competency-based curriculum was developed alongside an assessment framework to determine the minimum standard of learning in each grade (Indonesia #3181). For example, Lesotho began administering grade 3 assessments

[10] https://www.whitehouse.gov/fact-sheets/2025/04/fact-sheet-president-donald-j-trump-reforms-accreditation-to-strengthen-higher-education/ (accessed April 24, 2025)

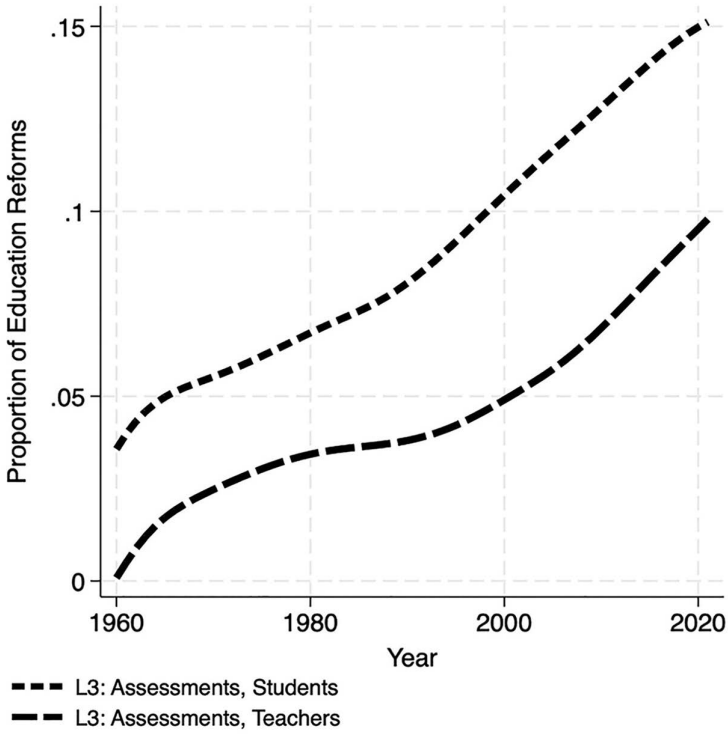

Figure 5.5 *Proportion of assessment reforms for students and teachers*

in reading, science and mathematics and determined that when the students scored above 70% in each subject, they achieved minimum competency (Lesotho #4521). These are just a few examples of reforms that aim to standardise minimum proficiency and assess student levels via assessments.

Two fundamentally important issues must be considered when understanding the rise in government efforts to standardise knowledge. First, who determines these standards, including the cutoff score that distinguishes students who are proficient from those who are not? Who creates these assessments and what metrics are used to determine the cutoff score? Knowledge is a social construct; those who hold authority shape and validate knowledge. The standardisation of knowledge can be used

to promote greater accountability to ensure that public spending on education is used in a way that adheres to a certain level of quality. This standardisation is also a way for the government to inform the public about its efforts to improve the education sector. Additionally, quality standards may be determined arbitrarily. What happens to students when they do not meet the minimum standards set by authorities? Are students asked to repeat the grade? Are they denied the opportunity to advance to the next grade level? These policy decisions place students – many of whom are denied further education because they do not meet the minimum standards – into various streams that profoundly impact their short- and long-term opportunities stemming from access to higher levels of learning.

Pathway for learning

The most common way to alter the pathway for learners is by introducing or abolishing existing assessments. In 2010, Malta eliminated examinations for 11-year-old students to reduce academic tracking into secondary education (Malta #4753). Many countries followed a similar path by eliminating the school leaving exam to expand access to secondary education against the backdrop of the World Declaration on Education for All.[11] In other cases, national examinations were offered to restrict students' advancement to the upper years of education. For example, in Eritrea, a reform introduced in 1995 changed national examinations by stipulating that these examinations would be conducted four times a year in the final year of middle school. The exam results were then used to determine who was admitted to upper secondary schools (Eritrea #2244). These policies were designed to control the increase in the number of students entering the education system. Examinations serve as a tool for stratifying society through the selection of a handful of learners who can advance to higher educational levels.

Through categorisation and sorting, this form of stratification creates greater inequality in societies (Domina et al., 2017). Schools have historically served to classify students into levels, thereby perpetuating existing inequality. The theory of categorical inequality refers to the organisational stratification of individuals who define and perpetuate

[11] World Declaration on Education For All, Jomtien, Thailand (1990) https://bice.org/app/uploads/2014/10/unesco_world_declaration_on_education_for_all_jomtien_thailand.pdf

social inequality (Tomaskovic-Devey et al., 2009). Assessment is a perfect example of categorical inequality because it categorises learners according to their test scores to determine their entry into higher levels of education. Sorting pupils is beneficial for countries in various ways. It serves as a mechanism that can be used to direct the flow of students from one level to another while adhering to the principles of equity and access that are pre-determined by the government. The public interprets these standards as impartial and objective, raising little concern over the validity of the results.

Competency-based learning

Student assessment reforms that mention competency-based evaluation began to appear more frequently in the mid-2000s. We observe more reforms that focus on competency-based learning and evaluation, including competencies that are identified by the OECD: knowledge, skills, attitudes and values (OECD, 2019). Zambia articulated a reform that aims to create a new competency-based curriculum emphasising the importance of "cognitive and non-cognitive skills development of youth" (Zambia #10823). This trend is also reflected in early childhood education; for example, Costa Rica prioritised competencies in "self-knowledge, interaction with the environment; socio-cultural interaction; expression, communication and representation" (Costa Rica #1563). This growing emphasis on competency-based learning and student evaluation reveals that more and more countries are considering alternative ways to assess student learning.

While several references to competence-based assessments have been made in recent years, the nature of these assessments needs to be further investigated. Competency can be defined in different ways, and its usage is not systematic across countries. Competency is often associated with twenty-first century skills such as critical thinking, creativity, collaboration and communication (4Cs).[12] Some reforms reflect this definition of competency and emphasise the importance of proficiency beyond content knowledge. In some cases, competency is defined as the mastery of knowledge in a specific domain through the demonstration of knowledge and skills, which resembles the use of formative assessments in classrooms. Formative assessments are leveraged as evaluative tools to assess

[12] For more information on the 4Cs, refer to: http://exploresel.gse.harvard.edu/frameworks/3 (accessed on April 1, 2025).

student learning with the aim of improving teacher instruction and content delivery, while summative assessments measure cumulative learning and are typically conducted towards the end of the unit or school year (Dixson & Worrell, 2016). Reforms related to formative and summative assessments have appeared more frequently in recent decades to evaluate a newly revised curriculum (Singapore #5534) and providing teachers with a tool to evaluate students' content knowledge (Denmark #1922). The analysis reveals the various definitions of what constitutes 'competency-based' education reform and presents increasingly varied ways of assessing students and their learning.

Assessments: Teachers

Policymakers have considered various policies to improve the effectiveness of teachers. The content of the reforms in this study was analysed using the three constructs of teacher effectiveness. The survey of assessment and teacher reforms yielded four types of reforms: 1) teacher qualifications, 2) evaluation of teacher performance, 3) pay for performance and teacher incentives, and 4) assessments that support teachers and their pedagogy.

Teacher qualification

The first category of teacher assessments is related to pre- and in-service teaching training and includes traditional methods and the assessment of teacher quality, such as the administration of an exam to measure competency after a teacher goes through in-service training provided by degree-granting higher education institutions. Most reforms focused on the evaluation of teachers in the 1960s and 1970s are related to this typology of teacher evaluation, especially with respect to the importance of in-service teacher training programmes (e.g., Chile #1262, Djibouti #1992 and Finland #2360). These reforms continue to be articulated in the 1990s and onward, making greater references to standards with the aim of maintaining the overall quality of teachers (e.g., Israel #3334 and Estonia #2285). In some cases, teacher qualification aims to equalise teacher quality between public and private schools in light of the rise of privatisation of education (Grenada #2807 and Singapore #5539).

Evaluation of teacher performance

Reforms related to teacher performance appeared for the first time in the 1990s and continue to be articulated in the early 2010s (e.g., Albania #7054, Mexico #19781, Georgia #2573, and Argentina #1011). These

reforms are typically referenced alongside other reforms that aim to assess student proficiency with reference to a specific assessment process to evaluate teacher quality. For example, the Syrian government stated the following:

> Each province chooses a subject and grade in which a standardised assessment is administered to all students. The main purposes of the provincial assessments are to identify students' learning levels, diagnose learning difficulties, evaluate teachers' performance and review the curriculum for the purpose of further development. (Syria #6307)

Similarly, in Egypt, a personalised portfolio was created for every teacher. These portfolios consisted of performance evaluation reports (including records of professional development training) that were completed by each teacher's supervisor, such as their head teachers or school principals (Egypt #2122). Oversight of the curriculum and teacher performance has increased in the country since the 1990s, with the public education system becoming more involved in regulating and evaluating how teachers teach in their classrooms. Turkey constitutes another example of this: in 2016, a new reform called the Teacher Induction Programme introduced a multipronged assessment process to assess teacher quality through classroom observations, assessments conducted by the school administrator/inspector and a written examination taken by the teacher (OECD, 2020). In Mexico, a new teacher appraisal reform was implemented in 2015 that resulted in the assessment of over 150,000 primary and secondary teachers. Teachers who scored lower than the average score obtained in teacher assessment were then required to undergo additional training (Mexico #4831).

Pay for performance and teacher incentives

Education reforms related to pay for performance began to appear in the early 2000s. In Colombia, a reform introduced the use of teacher evaluations, which are used to determine teachers and their career progression in education, including leadership roles (Colombia #1439). Several reforms linking teacher remuneration with evaluations have also been introduced in countries such as Ecuador (#10669). Italy also moved to merit-based teacher salaries by explicitly mentioning that teacher compensation is tied to the achievement of performance criteria that are determined by the evaluation committee at the school level (Italy #3381). In some cases,

such as in the Teacher Service Commission Act in Kenya, teachers with poor performance are asked to leave the teaching profession.

Assessments that support pedagogy

The last category of reforms refers to the provision of assessment and evaluation tools to improve pedagogy in classrooms. In Ethiopia, teachers began to take part in evaluations, ensuring that they play an influential role in the design and implementation of curricula (Ethiopia #2322). In the 2000s and 2010s, several reforms encouraged teachers to make use of student-level data to help inform their own teaching in the classroom (e.g., Jordan #3599, United States #6763, Canada #1162 and Israel #3343). Without adequate training and support for teachers, these assessments may increase teachers' burden. The actual implementation of reforms associated with the use of evaluation tools will be important for assessing how such reforms support teachers and their pedagogical approaches.

Trends in global patterns of teacher evaluation and assessment reforms reveal the growing pressure that teachers experience in the field of education. The traditional modality of teacher evaluation is no longer the defining factor in defining teacher quality. Teachers are held to different standards every time a new set of standards for teacher performance and evaluation is introduced. Teachers are also held accountable for a variety of aspects that affect the learners, including the transmission of content knowledge, curriculum design and implementation, and test scores. Teachers conduct day-to-day evaluations of students and their learning as they are also being evaluated by their peers and superiors, including other teachers, school principals, inspectors and professional evaluators. Teachers also need to adapt to the external changes that affect them, including funding cuts for education, the ability to cater to a diverse set of learners and other external conditions that affect teaching and schooling, such as natural disasters and the pandemic. The roles and responsibilities of teachers have grown increasingly complex as countries enact more assessment policies towards greater oversight and monitoring of teachers, who should be empowered to teach and guide students. The primary concern expressed by teacher unions is that the overemphasis on assessments has undermined teachers and their agency (Bascia & Osmond, 2013). This study's findings corroborate studies confirming that teachers are placed under extra pressure to promote greater accountability in education and that teachers are held accountable for students and their learning.

6. CONCLUSION

This study has several key takeaways. Overall, we observe an increase in the prevalence of assessment reforms over a period of 60 years. Second, we find evidence related to the prominence of education policies associated with assessments – this coincides with the proliferation of market-based neoliberal policies in education. Third, the typologies of assessment reforms have increased and shifted over time to place a greater emphasis on assessments that measure both student learning and teacher competency. Finally, we find that assessment reforms are influenced and affected by regional and international forces.

This study critically reveals the tendencies of governments to articulate education reforms to gain greater control and management of their education system. The rise in reforms related to standardisation of knowledge – evident from the rise in reforms that focus on standard setting and the development of data management in education, including large-scale assessments and quality assurance mechanisms – reveals the growing demand for regulation and systems that help control the production of knowledge.

Systems of evaluation help governments gain control over institutions, including schools, universities and academic institutions. Assessments and accreditation help determine what knowledge is produced and who should claim this knowledge. These systems of control serve as a mechanism to provide greater oversight and control by those in authority to regulate and monitor educational institutions and their recruitment, administrative and fiduciary responsibilities.

Additionally, the study points to how countries justify their policy choices by externalising reforms and benchmarking learning with international standards. Referencing these international benchmarks and international rankings enables governments to legitimise their policies, thereby allowing them to gain the political and financial support necessary to enact large-scale reforms. Countries engage in the rationalisation of their policy decisions by referencing other countries and their best practices.

This study highlights other related research topics, such as the determinants of assessment reforms, analyses of the sequencing of reforms, the magnitude of assessment reforms and the role of various stakeholders in formulating reforms. Further analyses of national discourses combining global and cross-national analysis and case studies will reveal how

political, economic and historical contexts influence the evolution of assessment reforms. Furthermore, further analysis of policies and practices associated with reforms will identify the interests of various players in education, including policymakers, researchers, school leaders, teachers and parents. These studies will determine the context for why and how reforms are articulated, providing a deeper understanding of how countries prioritise reforms to effect changes to their national education systems.

6. Accountability from above: examining international organisations in education policy

Kerstin Martens

1. INTRODUCTION[1]

Education and education policy have undergone major reform processes over the last few decades (Bromley et al., 2023). The international level of education policy-making as well as transnational processes have gained significance, fostering trajectories and translations for a global script for education. While education policy had long been regarded as a predominantly national domain with little need to adapt to international influences, this perception has changed over the last three decades. In the 1990s, the General Agreement on Trade in Services, led by the World Trade Organization, promoted transnational processes in the education service sector (Robertson et al., 2002; Verger, 2009). The turn of the millennium saw a proliferation of data generation in education across states, particularly through international large-scale assessments (ILSAs). At the same time, issues involving the harmonisation of education systems to allow for more mutual exchange across Europe, as well as in other world regions, required increasingly coordinated efforts across borders (Jungblut et al., 2023; Vögtle & Martens, 2014) to solve. Today, education

[1] This chapter is a product of the research conducted in the Collaborative Research Centre 1342 "Global Dynamics of Social Policy" at the University of Bremen. The centre is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 374666841 – SFB 1342. We thank Judith Albrecht and Duncan MacAulay for their assistance in conducting research for this paper and Dennis Niemann for his valuable comments on an earlier draft.

policy, like other policy fields, is a topic of international cooperation and concern (Niemann et al., 2023).

In this respect, international organisations (IOs) are often seen as highly valued sources for initiating and shaping national education reform processes. IOs are public international actors that exclusively have states as their members. At least three states must agree to establish an IO through an arrangement, such as a treaty or convention. Because of this condition, IOs can also be referred to as intergovernmental organisations. States establish IOs to bundle competencies in a particular policy field, exert more influence over third parties or to receive advice and guidance. As a result, IOs can hold states accountable through their initiatives for norm application, reform procedures or peer reviewing processes. The spectrum of IO competencies varies depending on their mandates: IOs may have overarching capabilities – as is the case for the European Union (EU) – or provide reviews for education reforms, evaluate best practices and collect comparative data to evaluate systems.

IOs' datafication initiatives have had a significant impact on national education policies. The governance-by-numbers or "soft governance by hard fact" approach (e.g., Grek, 2009; Niemann & Martens, 2018; Steiner-Khamsi et al., 2024a) has become a prominent tool of global governance used to exert influence on national governments. Unsurprisingly, the Organisation for Economic Cooperation and Development (OECD) and its Programme for International Student Assessment (PISA) have piqued the interest of many political scientists who have turned their attention to the education sector. PISA has drawn worldwide attention far beyond the core membership of the OECD. Although other ILSAs before PISA had been led by international agencies – most importantly the Trends in International Mathematics and Science Study (TIMSS), which was organised by the International Association for the Evaluation of Educational Achievement (IEA)[2] – PISA has received far more attention from the media and policy members due to its promotion, resources and rationales (Addey et al., 2017; Addey & Sellar, 2018).

This chapter explores how IOs have contributed to the current global script of data for accountability. It explores how the field of IOs active in education has developed since WWII. We adopt the perspective of sociological institutionalism (DiMaggio & Powell, 1983; Meyer et al., 1997)

[2] For details on the work of the IEA prior to PISA, refer to the work of Pizmony-Levy (2013).

to argue that IOs contribute to the global diffusion and transfer of policy ideas. In previous studies, we identified and explored a set of 30 IOs[3] that were actors in education policy (Niemann et al., 2023; Niemann & Martens, 2021). Our analyses revealed that while education has always been a relevant topic for some IOs due to their mandate, many others have adopted an educational agenda over the course of their existence and for varying reasons (Martens et al., 2024). By leveraging approaches to IO institutional design from international relations (Debre & Dijkstra, 2021; Hooghe et al., 2019), common characteristics and comparative analyses with other fields are applied to these 30 IOs. The analysis spans from their global aspects to their regional ones as well as from their cultural features to economic aspects. Perhaps the most surprising empirical finding is that two-thirds of all IOs active in education are regional IOs with membership bound to states geographically based in only one area of the world.

Moreover, we examine the mechanisms by which these 30 IOs contribute to the global education script(s) and how they distribute their

[3] ABEGS: Arab Bureau of Education for the Gulf States, ADB: Asian Development Bank, AfDB: African Development Bank, ALECSO: Arab League Educational, Cultural and Scientific Organization, APEC: Asia-Pacific Economic Cooperation, ASEAN: Association of Southeast Asian Nations, ASEF: Asia-Europe Foundation, AU: African Union, CARICOM: Caribbean Community, CBSS: Council of Baltic Sea States, CW: Commonwealth of Nations, ECCAS: Economic Community of Central African States, EFTA: European Free Trade Association, EU: European Union, IADB: Inter-American Development Bank, IFESCCO: The Intergovernmental Foundation for Educational, Scientific and Cultural Cooperation, ILO: International Labour Organization, ICESCO: Islamic World Educational, Scientific and Cultural Organization, Mercosur: Southern Common Market, OAS: Organization of American States, OECD: Organisation for Economic Co-operation and Development, OECS: Organisation of Eastern Caribbean States, OEI: Organization of Ibero-American States for Education, Science and Culture, SAARC: South Asian Association for Regional Cooperation, SEAMEO: Southeast Asian Ministers of Education Organization, UNASUR: Union of South American Nations, UNESCO: United Nations Educational, Scientific and Cultural Organization, UNHCR: United Nations High Commissioner for Refugees, UNICEF: United Nations International Children's Emergency Fund, WB: World Bank.

data, beliefs and recommendations, through which they can hold nation states accountable. Drawing on the categorisations proposed by Nagel et al. (2010), Martens and Jakobi (2010) and Armingeon (2021), we outline the diverse channels IOs use to influence policy processes and then assess the extent to which the sample of 30 IOs practically apply these channels. We reveal how mechanisms of influence vary across IOs with a global and regional focus, as well as between economically oriented and humanistic IOs. Unsurprisingly, while (almost) all IOs seek to influence national policies by producing and distributing norms and standards through policy briefs and documents, only a few have "hard" mechanisms of funding and conventions at their disposal. Moreover, PISA – the OECD's primary tool for collecting and evaluating education – stands out as a globally accepted and applied standard for the datafication of education.

We take three steps in our theoretical approach: sociological institutionalism helps us analyse the fields of IOs contributing to the proliferation of a global script in education, while the approaches of institutional design explain the differences between IOs. Adding mechanisms of governance enables a differentiation between the channels of influence used by IOs to hold states accountable. We draw on a wide range of literature from different fields – sociology of education, political science and international relations, public policy and education studies – to develop our categories. Our empirical analysis assesses how the field of education IOs is constituted, how it can be decomposed and how IOs exert influence. This is the study's first empirical snapshot; thus, its main purpose is to identify patterns rather than to paint a comprehensive picture.

The chapter is structured as follows. We first present the theoretical framing of this paper to outline how the field of IOs active in education is constituted, how it can be characterised and how IOs exert influence. We then contextualise research about education IOs and provide a methodological approach for determining what can be categorised as an "education IO". Following this contextualisation, we present how the field of education IOs is constituted and how it can be characterised using the framing of the theoretical section. For the purposes of this chapter, the empirical analysis of mechanisms focuses on standards such as the Sustainable Development Goals (SDGs) and on international large-scale assessments, such as PISA, as these two issues dominate our empirical analysis. Stated briefly, our analysis finds that Western globally widespread conceptualisations prevail despite a predominantly regionally structured education sphere.

2. THEORETICAL APPROACH: GRASPING THE REALM OF EDUCATION IOs

Sociological institutionalism focuses on the ways in which institutions shape and influence social behaviour, interactions and outcomes. This concept emphasises the roles of both formal and informal rules, norms and cultural values in shaping individual and collective behaviour within social systems (Meyer & Rowan, 1977). When seen from this perspective, institutions are not just passive structures but dynamic forces that both reflect and shape societal norms and practices. When the perspective of sociological institutionalism is considered, IOs can be deemed crucial in shaping and influencing social behaviour and outcomes on a global scale. They establish formal rules, set norms and formulate standards that govern interactions among nations and transnational actors. Sociological institutionalists argue that IOs act as powerful influencers of the behaviour of states and other actors by providing ideational frameworks for international cooperation, conflict resolution and resource allocation. Furthermore, IOs contribute to the diffusion of ideas, values and practices across borders, thereby influencing societal norms and behaviours in diverse cultural contexts. Through their policies, programmes and initiatives, IOs can impact various aspects of society.

Sociological institutionalists analyse organisations by examining how they emerge, evolve and exert influence (Meyer & Bromley, 2013). By understanding the dynamics of and in IOs, sociological institutionalists gain insight into the complexities of global governance and the mechanisms through which global social order is constructed and maintained (Finnemore & Sikkink, 1998). Overall, the role of IOs in sociological institutionalism highlights their significance in shaping the global social order and influencing the behaviour of states and other actors in the international arena. Therefore, while sociological institutionalism provides the basis for our argument by highlighting how IOs spread norms and values across the globe, it does not differentiate between different types of IOs and their specific characteristics, which shape the scope of their activities. Differences may also arise at the structural level of IOs. These differences are what scholars of international relations (IR) refer to as IOs' institutional designs and they allow for distinct analyses of the organisational fields of IOs that are active in education.

Institutional design approaches to IOs provide a framework for analysing the composition of an organisational field. These approaches allow

for a differentiation between IOs active in education policy. Building on earlier work by Rittberger et al. (2019), other scholars explored how institutional factors, which often originate from the IO's founding principles, influence the spectrum of how IOs approach policies (Hooghe et al., 2019; Lenz et al., 2015). Approaches to institutional design resonate with rational choice approaches and assume that the setup influences the scope of activity that IOs can exert (Hawkins et al., 2006; Nielson & Tierney, 2003). Components of institutional designs are therefore essential in understanding what IOs (can) do and why they do it. These components may also determine IOs' mechanisms of governance used to influence states and state behaviour and hold these states accountable. Institutional designs also determine how autonomous IOs expand their spectrum of activities and act in organisational fields (Niemann et al., 2021, p. 14).

The scope of the issues IOs cover and the membership rules they adhere to constitute two important features of institutional design, which shape IO activities (Koremenos et al., 2001). Two notions regarding the scope of IOs are commonly distinguished: comprehensive responsibility and the policy-specific concerns of IOs (Rittberger et al., 2019). Thus, IOs with comprehensive responsibilities (e.g., the UN or the EU) deal with a variety of policy issue areas, whereas policy-specific IOs, such as the United Nations Educational, Scientific and Cultural Organization (UNESCO), the Organization of the Petroleum Exporting Countries or the International Migration Organization have a limited-scope spectrum. Accordingly, comprehensive IOs are better able to link educational issues with other topics in their portfolio (e.g., economics, security issues or climate change) (Niemann et al., 2021). IOs are typically characterised as having either universal or restricted membership (Rittberger et al., 2019). Some IOs allow all states to potentially become members, while other IOs limit membership to states with specific characteristics, such as geographic belonging, religious dominance or cultural heritage. Homogeneity in IO membership ultimately makes decision-making processes easier.

How do IOs influence national policies? Impact studies often have difficulties both conceptualising the differences and empirically evaluating causality. IR deals with theories with IOs as actors as well as their influence on states from a constructivist perspective (Barnett & Finnemore, 2004). In this sense, IOs' "actorhood" is assumed; they are not seen merely as state instruments but rather as entities that enjoy a certain level of independence from their principals. However, IOs are

not unitary actors that speak with one institutional voice. IOs consist of different units, subsections or departments that may have different mechanisms or foci in their activities depending on the power relations or hierarchies between them (Armingeon, 2021). In a similar vein, IOs are not static; their spectrum of activities can change over time and often expand across policy fields. Furthermore, education policy is no exception when it comes to IO missions broaching other fields (Hall, 2016; Niemann et al., 2023).

The literature contains studies that conceptualise the different mechanisms by which IOs exert influence and hold states accountable. Early work by Jacobson (1979), for example, distinguishes between five categories of IOs' work, such as informational activities, normative activities, rule-creating activities, rule-supervisory activities and operational activities. McNeely and Cha (1994) distinguished between exchange of information, charters and constitutions, standard-setting instruments and technical and financial resources, whereas Armingeon (2021) identified "channels of influence", namely the resources, constraints, standards, evaluations and information used by IOs to exert influence on national policies. In our work on the OECD, we differentiated between "mechanisms of governance", specifically idea generation, policy evaluation and data production (Martens & Jakobi, 2010). In addition, we argue that the exercise of mechanisms and their choice needs to be analysed in context, specifically in terms of their conditions and effects. Our prior work (Nagel et al., 2010) differentiated between five "IO governance instruments", namely, norm-setting, opinion formation, financial means, coordinative activities and consulting services (e.g., the Bologna Process) for analysing the European agenda to recognise higher education policies across the continent.

These studies all view IO instruments of accountability (in education policy) in similar ways. They highlight IOs' informal and informational character in addition to the coercive powers they may have. Summarising and drawing from these works, we typologise and differentiate IO mechanisms to study the set of 30 IOs operating in education policy. We focus on the characteristics of these mechanisms to determine which of them are regularly employed by IOs active in education policy. We distinguish between coercive mechanisms (such as conventions and financial means), standards-setting mechanisms (such as policy recommendations and idea production) and coordinative activities (such as evaluations and consulting purposes). This typification allows us to distinguish the ways in which IOs operate on the international level; at

the same time, the types can be applied to create workable categorisation of IO mechanisms. This is different from domestic contexts in that any implementation of IO outputs requires another step – namely, a transfer to the national level – before they can be put into practice (Steinebach et al., 2024).

2.1 Coercive Power through Conventions and Financial Means

IOs active in education policy can generally be expected to have little means of exerting coercive power. Unlike other policy fields of "hard politics", such as security policy or economic policy, martial power or economic sanctions do not play a role in education policy. However, IOs have two ways of directly influencing countries' education policy: they may give financial resources to countries to support local education projects (e.g., the building of education infrastructure) or they may set up binding conventions to which countries are obliged to adhere. IO expenditures are often intended to support specifically defined projects; therefore, they are bound by conditionalities. While conventions do not involve financial issues, they have a binding character for signatory countries. Although IOs often lack the means to enforce noncompliance, they can "name and shame" countries that do not adhere to their previous agreements and hold them accountable. When seen from a social institutionalism perspective, conventions that provide global standards are issued by globally active IOs, whereas financial means for education are provided by IOs with the distinct institutional design of lending money; thus, any IO with this institutional design to be able to work with money brings "banks" to mind. Armingeon (2021), however, expected financial resources as an IO mechanism to generally be of minor relevance, since only a few IOs have access to these resources.

2.2 Standard Setting through Policy Recommendations and Idea Production

IOs regularly produce policy papers and publish briefs containing recommendations for their member states. This mechanism aims to steer policy without being binding, as IOs seek to provide guiding principles. "IOs set standards against which national policies are evaluated and which create normative pressures for the national context of policymaking" (Nagel et al., 2010, p. 10). For example, the OECD may suggest increasing spending on education or recommend improving the level of

graduates to meet the requirements of the job market. This tool is often described as "soft" governance, as the mechanism does not contain any coercive power. Others have referred to such activities as knowledge brokerage (Niemann & Martens, 2018). IOs leveraging such a mechanism find problems and deliver solutions based on knowledge production. However, IO activities in the form of idea development, knowledge production and opinion formation should not be underestimated when it comes to their impact on governments, as they shape the way countries and representatives of countries think about an issue, what they consider to be important for future policy proposals and what solutions to problems are available. In this way, IOs lead debate, create consensus and provide tacit approval (Armingeon, 2021). This is presumably the main mechanism for IOs to exert influence; thus, almost all IOs, regardless of their institutional design, can be said to produce norm- and standard-setting documents to influence the behaviour of states.

2.3 Coordinative Activities through Evaluations and Consulting

In addition to binding conventional financial means and standard-setting mechanisms, IOs generate, gather and publish comparative data in education policy. IOs are the main and sometimes only sources of information for these states, especially in terms of comparative analyses and evaluations across nation states. IOs are in fact the prime actors for some of these evaluations, as they oversee the process, set goals and criteria, and organise analyses and publications (Martens, 2007). For other evaluations, IOs gather information from governments and governmental subunits. Data from IOs are usually considered to be objective; this allows for direct and indirect comparisons and evaluations of states to be made over time. By providing such data, IOs directly and indirectly influence the datafication of education and political debate, which is often conducted through the media, for which such comparative analyses are easily digestible. In fact, Armingeon (2021) considered this mode a major means of influencing national governmental policies. Any evaluation and consulting may lead to global convergence and what sociological institutionalists call "world culture". However, it is also possible that in more regionally active IOs, adherence to any contextualised results takes precedence over world values due to increased homogeneity amongst member states and the prevalence of cultural factors that are distinct to geographical areas or otherwise bind member states.

IOs' mechanisms have been characterised in various ways. Elfert and Ydesen (2023) stated the following regarding the three largest organisations active in education, namely UNESCO, World Bank and the OECD: "We will conceptualise UNESCO as the 'idealist,' the OECD as the 'master of persuasion,' and the World Bank as the 'master of coercion'" (p. 25). In sum, drawing from research on sociological institutionalism, institutional design and approaches to IO mechanisms allows us to grasp the field of IOs active in education and categorise them according to their characteristics. This enables us to evaluate the ways in which they exert influence on states. We then assess the resulting theoretical framework using empirical data. We first describe how the field of education IOs developed over time before examining the mechanisms by which they exert influence. Lastly, we evaluate the extent to which IOs in education contribute to the datafication of education and how they hold states accountable.

3. ANALYSING IOs IN EDUCATION POLICY: GROWTH, DISTRIBUTION AND MECHANISMS

The political science discipline discovered the field of education (Busemeyer & Trampusch, 2011; Jakobi et al., 2010), only with a time lag. However, scholars have gradually caught up in their examinations of the determinants of national reform processes, which has allowed them to compare skill formation across borders and explore the historical or political paths of education policies throughout the world. Given the blurring boundaries between sociology and education science and, to a limited extent, comparative economics, scholars in these research areas have also started increasingly exploring the contribution of IOs in these fields. In particular, the OECD, UNESCO and the World Bank triggered many studies. However, despite this growing trend, Kijima and Lipscy (2024) proclaimed that from the perspective of IR, education is still wrongfully overlooked as a field of interest: "Although international education is an often-neglected topic among international relations scholars, the substantive importance of education is unmistakable: on average, governments devote about 14% of their spending on education, about double the amount they invest in military expenditures" (p. 2).

This ignorance of IR in the realm of education policy is surprising, since IOs in particular have been recognised as important actors. IOs in education exert autonomy and authority simultaneously (Niemann, 2012) in the same way as IOs in other fields. They provide international

education policy data, develop normative guidelines that become globally applied and provide financial support for education projects across the world. In doing so, IOs contribute to the datafication of global education and education policy (Jarke & Breiter, 2019). At the same time, national education systems are under pressure to respond to the challenges posed by globalisation processes. These systems must maintain national economic competitiveness in world markets and keep pace in a constantly growing, knowledge-based economy that sees education as the key to growth. Moreover, these systems need to prepare their graduates for global competition in labour markets.

Education and IOs have long been the subject of discussion in academic works, primarily those written by educationists and sociologists. However, aside from case studies on single IOs or comparative approaches involving a few IOs, little has been revealed about education IOs until recently. Most prior studies dealt with the "usual suspects": UNESCO (e.g., Lerch & Buckner, 2018; Mundy, 1999; Rittberger, 1995; Jones, 1988; Krill de Capello, 1970), the World Bank (e.g., Jones, 2007; Mundy & Menashy, 2014; Mundy & Verger, 2015) and the OECD (Grek, 2014; Niemann & Martens, 2021). Notable exceptions include works by McNeely (1995), Mundy (2007), Resnik (2006) and Moutsios (2009), which dealt with education IOs in a comparative perspective; these studies, however, did not examine how the field as a whole is constituted. Thus, the extent to which other IOs deal with education and education policy has not been examined (or if so, only rarely). If and how education as a policy field has spread across and within IOs has become a subject of interest for us in terms of developing a more holistic picture of how education is dealt with on a global scale (Martens et al., 2024).

We were also interested in exploring how the field of education IOs can be characterised. To identify the field of education IOs, it is necessary to specify what counts as an IO and how an education IO can be defined. For this work, we (Martens et al., 2024; Niemann & Martens, 2021) determined that we would first limit our sample to "public" organisations, namely IOs in which states are the prime members. This limitation, of course, leaves out NGOs such as the Melinda and Bill Gates Foundation and the Soros Foundation as well as think tanks such as the Global Partnership for Education, which can be financially strong and important players in the global education sphere. Certainly, some of the NGOs and think tanks active in the education sphere are important for shaping and providing education and education policy (Boli & Thomas, 1999; Bromley et al., 2023); however, analysing them and their

characteristics would increase the study's complexity and would require further exploration. Therefore, for the purposes of this work, we kept trained analysis on one clearly identifiable group of actors: intergovernmental actors (Martens et al., 2024).

Then, established datasets of IOs from IR were taken into account in order to identify the ones that deal with education. Based on the Yearbook of International Organizations) as well as the Correlates of War dataset, all active IOs as of 2020 were reviewed to determine whether they pursued education policy as a field of activity. We considered all formal levels of the education sector and did not differentiate between IOs active at the K-12 level (primary and secondary grades) and/or the tertiary education level (higher education and vocational training) (Martens et al., 2024). IOs were considered education IOs if they possessed three complementary features: (1) education was mentioned in the IO's programmatic mission statement as a designated task of the IO, such as in the IO's preamble, its (amended) treaties or its current web presence; (2) the IO had its own permanent organisational subdepartment, unit or otherwise named structural component that specifically dealt with issues of education or training (it is insufficient to have only one person dealing with education); and (3) the IO addressed education policy issues. Therefore, we do not include any organisations that focus on educational matters, such as teaching methods or facilitating scientific collaboration, in our definition. For example, while the North Atlantic Treaty Organization engages in educational activities, these activities are primarily aimed at training the organisation's own personnel in security-related areas (Martens et al., 2024, p. 484).

Using these toolkits and reviewing each IO individually, a total of 30 organisations were identified. This sample forms, in our view, the population of active education IOs (Niemann & Martens, 2021). The field of education IOs expanded steadily over time; only two organisations were active in 1945, whereas 30 organisations were active in 2020, as shown in Figure 6.1. Unsurprisingly, UNESCO and the International Labour Organization (ILO) were the first two IOs to work in the field of education policy after WWII, followed by the Organization of American States (OAS), the Southeast Asian Ministers of Education Organisation (SEAMEO), the Organization of Ibero-American States for Education, Science and Culture (OEI), and the Commonwealth and other UN-related institutions, such as the United Nations High Commissioner for Refugees (UNHCR), the World Bank and the United Nations Children's Fund (UNICEF) until 1965. By the mid-1990s, this

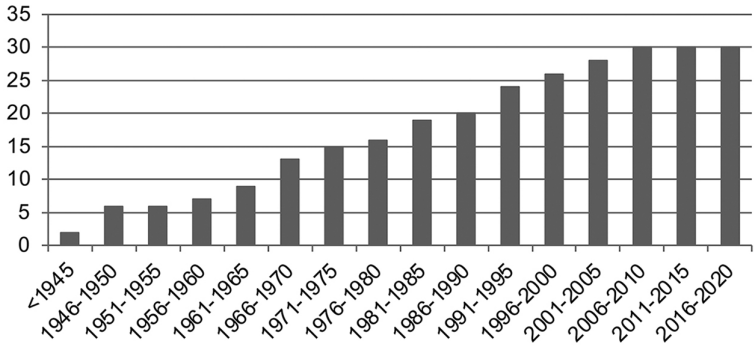

Figure 6.1 The growth of education IOs

growth had slowed, and only two new education IOs entered the field after 2005: the Union of South American Nations (UNASUR) and the Intergovernmental Foundation for Educational, Scientific and Cultural Cooperation (IFESCCO). Education IOs are now working in all corners of the world. Thus, it is safe to say that the field of education IOs has expanded steadily beyond primarily Western-based values across the world – as sociological institutionalism would assume. It also seems that the field of education IOs became saturated around 20 years ago.

This field of 30 IOs can be further analysed by focusing on the institutional design of IOs and their accompanying intrinsic characteristics. In a nutshell, education IOs differ in their geographical reach of operations as well as in their thematic scope. However, these features are not mutually exclusive, and IOs may overlap in both categories. We identified three different types of education IOs in terms of their geographic reach: global, regional and transregional (Niemann et al., 2023). Our three usual – UNESCO, the World Bank and OECD – are global education IOs. While it may be obvious that UNESCO and the World Bank are global players, as their membership is not restricted, we also classified the OECD, which has restricted membership, into this group, as its work in education is not exclusively provided to member countries. For example, more than 100 countries have participated in PISA since 2000.[4] Other globally active IOs involve organisations that fall under the UN

[4] https://www.oecd.org/pisa/ (last accessed: September 14, 2024).

system, such as UNHCR, UNICEF and the ILO, shows the geographic distribution (global, transregional and regional) as well as the scope of issues covered by education IOs (i.e., multipurpose, single-purpose, education or economic issues).

Most IOs active in education are, in fact, regionally or transregionally operating IOs whose membership is restricted mainly to states with specific geographic characteristics. Today, 20 IOs working in education are regional, and most of these IOs started working on education long after 1965. Regional IOs significantly shape modern international education policy, widening the organisational field and occupying niches. This group includes, for example, the Organisation of Eastern Caribbean States (OECS), the Council of Baltic Sea States, SEAMEO, the Arab Bureau of Education for the Gulf States (ABEGS) and the Economic Community of Central African States (ECCAS). Transregional organisations are characterised by member states from different regions and sharing commonalities, such as a religious focus, as with the Islamic World Education, Science and Culture Organisation (ICESCO), in which membership is connected to Islam, and the OEI, which is composed of states with Spanish or Portuguese backgrounds.

Regarding the thematic scope encompassed by IOs, we found seven IOs in our sample that promote education as a primary activity through their mandate. These IOs were founded with a specific purpose of focusing mainly on education policy as a major – and in some cases, solitary – mission. This group includes SEAMEO, ABEGS, the Arab League Educational, Cultural and Scientific Organisation (ALECSO), IFESCCO, the OEI and ICESCO. These IOs are all regionally focused; the only globally active IO in this group is UNESCO. All other education IOs, however, cover several policy fields – education is only one among many. On the one hand, the field is constituted by so-called "multipurpose" IOs (i.e., those with a broad spectrum and scope of issue areas). Examples of such IOs include the EU, the South Asian Association for Regional Cooperation (SAARC), the African Union (AU) and UNASUR. On the other hand, another group of education IOs are "specialised" or single-purpose IOs, which have a primary mission other than education, such as the UNHCR or the ILO, but encompass education policy in addition to their primary issue. For example, organisations such as the Asian Development Bank (ADB) or the African Development Bank (AfDB), which were set up with a specific mission in economic development, have gradually expanded their work into the education field.

Source: Martens et al., 2024, p. 485.

Figure 6.2 The world of education IOs

IOs with a prime mandate in economic policy have become an important subgroup in the organisational field of education policy. Almost one-third of today's education IOs constitute this group. Several "banks" started to include education in their portfolios at some point in time. Next to the World Bank belong the ADB, the Inter-American Development Bank and the AfDB in this category, among others. Furthermore, these economically oriented IOs substantially shape the global education discourse on education towards a human capital approach. Unlike other IOs, economic IOs emphasise how education improvements can be an investment in a country's future economic prosperity (Martens et al., 2024). Elfert and Ydesen (2023) pointed out that "While UNESCO's ontology can be traced back to the pre-World War II internationalist movement, both the OECD's and the World Bank's ideology is rooted in free-market capitalism and derives from the intellectual and political climate of the Cold War" (p. 24; similarly: Niemann, 2021).

4. MECHANISMS OF EDUCATION IOs

The mechanisms that education IOs apply to exert influence on both world politics and national policies vary across scopes and geographic locations. As expected, several IOs out of our set of 30 are able to exercise coercive power directly through conventions and financial means. In fact, the World Bank, ADB, AfDB and the AU are direct funders of education projects; thus, IOs that possess an economic scope, namely the IOs that function as banks, can direct national projects through this mechanism. Moreover, these IOs can hold states accountable for how they have spent the resources allocated to them. In addition, UNESCO gives out funds in times of emergencies, such as armed conflict or natural disasters. With the help of donors such as Education Cannot Wait and the Global Partnership for Education as well as country donors like Sweden and Japan, UNESCO establishes and carries out emergency education programmes on the ground.[5] Most IOs, however, have no set funding that they can use for education. Few IOs give out study grants. In this respect, the EU is a major funder of research and study. Unsurprisingly, ERASMUS+ has made 14 billion euros' worth of funding available to European universities through national and regional funding schemes.

[5] https://www.unesco.org/en/emergencies/education/need-know#how-does-unesco-support-countries (last accessed: September 14, 2024).

The EU also partners with the Association of Southeast Asian Nations (ASEAN), for example, in SHARE (Support to Higher Education in the ASEAN Region), a programme whose funding exceeded 10 billion euros between 2014 and 2019, including money provided to students participating at 32 different universities in the programme (e.g., through student mobility scholarships).[6] Some (trans)regional IOs, such as ICESCO, OAS and SAARC, give out research and study grants directly to individuals to encourage cultural cooperation and academic exchange. For example, in March 2024, ICESCO provided 200,000 US dollars in scholarships to 500 Palestinian students studying at Egyptian universities.[7] Other IOs, such as the Commonwealth, simply encourage member states to have an adequate portion of the national budget set aside for education. Overall, coercive power exerted via financial means is limited for education IOs. Most education IOs have no direct or indirect access to financial ways of supporting education.

The establishment of binding conventions and ratification by states is even less common for education IOs. The most prominent education initiatives led or supported by IOs, the OECD's PISA study and the EU-supported Bologna process, are purely voluntary programmes and are not based on any binding conventions. States can decide to leave these programmes at any time. Although some examples of conventions supported and executed by education IOs exist, they are mainly given out by global IOs and the EU. Examples of global conventions that are not explicit education conventions but deal with education as one significant aspect include the Universal Declaration of Human Rights of 1948 (which also proclaims the right to education), the Convention on the Rights of the Child of 1990 (which entails the full range of civil, political, economic, social and cultural rights, including access to education for children) or the United Nations Convention on the Rights of Persons with Disabilities, which recognises the right to education without discrimination on the basis of equal opportunities, thereby ensuring inclusive education.

[6] https://www.eeas.europa.eu/eeas/eu-support-higher-education-asean-region-eu-share_en (last accessed: September 14, 2024).

[7] https://icesco.org/en/2024/03/28/financial-support-from-icesco-for-500-palestinian-students-at-egyptian-universities/ (last accessed: September 14, 2024).

A major convention in the field of education is the UNESCO Convention against Discrimination in Education of 1960,[8] which became the first legally binding international instrument entirely dedicated to the right to education. It is also a cornerstone of the Education 2030 Agenda and seeks to advance inclusive, equitable and high-quality education for all. Like the Universal Declaration of Human Rights, it affirms that education is a fundamental human right. Furthermore, it obligates states to offer free compulsory education and bans any discrimination, thereby promoting equality and equal opportunities for education. However, only 110 states have signed the convention thus far.[9] In 2000, 164 participants of the UNESCO-led World Education Forum signed the Education for All commitment, making this initiative a global movement for providing quality basic education for all children, youth and adults and reaffirming the vision of the World Declaration on Education for All adopted 10 years prior (UNESCO, 2000). The EU has specifically focused on education and training and supports its member states in this respect in accordance with Articles 165 and 166 of the Treaty on the Functioning of the European Union. In response to the COVID-19 pandemic and extended school closures, the EU also recognised the need to address the challenges and opportunities of digital education with its 2021–2027 Digital Education Action Plan, which calls for the sustainable and effective adaptation of member states' education and training systems to the digital age (European Commission, 2021; Zancajo et al., 2022, p. 119).

Whereas coercive mechanisms such as funding and conventions are only exercised or applied by a handful of IOs, standard setting through policy recommendation and the production of ideas are mechanisms that all IOs can apply. It does not take much for IOs to execute these mechanisms and institutional features. In fact, it may even be argued that this is the raison d'être of many IOs: States set up global or regional IOs for the purpose of sharing and disseminating (common) standards and ideas. In this respect, IOs function as think tanks and distributors of shared values. Accordingly, with two exceptions, all 30 IOs in the sample apply these mechanisms. IOs typically produce policy documents, short policy briefs, press releases, and longer or shorter (academic) studies and the

[8] https://www.unesco.org/en/legal-affairs/convention-against-discrimination-education (last accessed: September 14, 2024).

[9] https://www.unesco.org/en/legal-affairs/convention-against-discrimination-education#item-1 (last accessed: September 14, 2024).

like. Many IOs have produced these documents periodically since their founding. However, two IOs seem to be rather inactive in this regard: the Russia-based IFESCCO and the Africa-based ECCAS. A recent study (Eilstrup-Sangiovanni, 2020) showed that IOs can die; if they still exist and show minimal levels of activity, they continue as "zombies" (Gray, 2018). Accordingly, the vitality can vary between IOs, between time periods and perhaps between policy fields. In our case, two regional education IOs and the two specialised IOs seem to be zombies, whereas global IOs with broad spectra seem to be more stable and exhibit more activity.

One particular global standard stands out in this regard: SDG 4. This has become the major reference point for a guiding standard on which many IOs base their policy goals for their member states. In a nutshell, SDG 4 requires states to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". IOs that are related to the United Nations, such as UNHCR, UNICEF and UNESCO as well as SAARC, ASEF, ADB and OECD, refer to SDG 4 when it comes to standards for education. Seen from the perspective of sociological institutionalism, SDG 4 is likely to facilitate a Western-based spread of norms through global IOs as these norms are adopted by regional organisations and passed on to member states. However, some IOs have developed their own strategic instruments that are more tailored to their specific regional needs, such as the OECS Education Sector Strategy (OESS) 2012–2016, which places improving education quality at the forefront of its policy. It emphasises collaboration and consultation between the OECS and its member states. According to the OESS, the plan "will be used by Member States to align their national Strategies and Plans".[10] As a regional organisation, OECS seeks to consider the philosophy and values of education in the Caribbean.[11] Similarly, the Asia-Pacific Economic Cooperation (APEC) developed its APEC Education Strategy 2016–2030, which is "the first education blueprint since the inception of the APEC"[12] and stressed the economic aspects

[10] https://drive.google.com/file/d/1SYg5EIX8FTvlACRKylw9Ml09sCUx_mN0/view (last accessed: September 14, 2024).

[11] https://oecs.int/en/our-work/knowledge/library/education/oecs-education-strategy (last accessed: September 23, 2024).

[12] https://mddb.apec.org/Documents/2020/HRDWG/EDNET-PREP/20_hrdwg_ednet_prep_005.pdf (last accessed: September 23, 2024).

of highly skilled citizens. Its vision is to "support sustainable economic growth, social wellbeing and employability of men and women in APEC economies".[13]

International large-scale assessments are prominent mechanisms of datafication. Not only do IOs increasingly use them as a tool for benchmarking and evaluation, but the "comparative turn" (Martens, 2007) to quantifiable assessments across states provides IOs with significant possibilities for influence. Nearly 80 states routinely participate in these assessments (Kijima & Lipscy, 2024, p. 4), which are operated by IOs. TIMSS and PISA are the best known of these assessments and typically evaluate major subjects, such as mathematics, sciences, and reading. The World Bank and UNESCO are the main providers of data in addition to these two large ILSAs; these data may be used as benchmarks. For many countries, depending on external funding, this mechanism exerts direct accountability on states when meeting these benchmarks is essential for receiving further aid (Besche-Truthe & Seitzer, 2025). Moreover, several new ILSAs focusing on regional developments in education have been initiated (Besche-Truthe & Seitzer, 2025, table 1). However, half of these are linked to the IEA or the OECD, which serve as conductors.

Overall, most education IOs do not set up their own specific ILSAs; rather, they use and apply OECD data, TIMSS or other datasets to analyse how their member states do in terms of education. Even if not all member states participate, some IOs make references to PISA for formulating recommendations for education policy. For example, ASEAN, UNHCR, UNASUR, ISESCO, IADB, European Free Trade Association (EFTA) or SEAMEO use PISA scores. ASEAN's Policy Brief on Safe School Reopening, Learning Recovery and Continuity directly referenced PISA 2018 scores for countries in East Asia and the Pacific. Another unexpected example of PISA's influence can be found in ICESCO. In May 2024, ICESCO, in partnership with the OECD and the Ministry of Science and Education of the Republic of Azerbaijan, held ICESCO's Second Ministerial Conference on PISA. As the name suggests, the conference discussed the PISA 2022 results among ICESCO member states, how states can improve student PISA scores and how ICESCO and the OECD could better support education reform among member states. Moreover, the Inter-American Development Bank (IADB) published

[13] https://mddb.apec.org/Documents/2020/HRDWG/EDNET-PREP/20_hrdwg_ednet_prep_005.pdf (last accessed: September 23, 2024).

PISA results on its website in 2012. Not only does the IADB publish the assessment results in both English and Spanish, but it also breaks these results down and publishes articles and studies on specific topics from the results, such as STEM results, short- and long-term trends and PISA participation numbers.

Among the 30 IOs in the study, those that did leverage PISA data mostly used it as a reference point either to inform their own policy recommendations or education framework. These IOs include ALECSO, ASEAN, the Commonwealth, EFTA, IADB, ILO, ICESCO, Mercosur, OAS, SEAMEO, UNASUR and UNHCR. As we know from previous studies, states that do not participate in PISA are also influenced by datafication through PISA results, as they use the data and findings of this ILSA to reform their education systems (Niemann & Martens, 2018, p. 273). SEAMEO runs its own regional learning assessment for Southeast Asian countries to improve learning outcomes for students in basic education: the Southeast Asia Primary Learning Metrics (SEA-PLM). APEC is planning to set up its own evaluation in 2030. From the perspective of sociological institutionalism, this usage of PISA as the main global provider of education data could be clearly seen as a sign of world culture inspired by Western values. As Besche-Truthe and Seitzer (2025) argue, "Most ILSAs were developed in and for the Global North; often with a Eurocentric and neoliberal perspective on education" (p. 11). Moreover, these Western ILSAs and their results are used by regional education IOs and Global South–based IOs for analyses. A Western model for evaluation is still clearly universally applied in this sense.

With very few exceptions (the two "zombies", IFESCCO and ECCAS, as well as SAARC and UNASUR), almost all IOs out of our set of 30 hold some sort of consultative meetings with senior education officials of their member states or even the ministers of education (from what can be found on their websites). Some also hold conferences on more specific topics. The sequence of meetings and the variety of participants involved differ. While at least one annual meeting is scheduled by every IO, others meet frequently on a variety of specialised topics. Our research shows that the three regional IOs – ASEF, EU and SEAMEO – hold the most meetings. IOs specialised in education, such as ABEGS, ALECSO and SEAMEO, meet on a wide variety of issues, including meetings or conferences on early childhood education, vocational education and adult education. In terms of consultations, IOs offer their framework of various issues for countries to follow for their own national policy. They

often present themselves as intermediaries for countries to meet and discuss various topics and issues.

5. CONCLUSION

In a nutshell, the world of education IOs has continuously grown since World War II. As in other policy fields, new IOs active in education were founded, and existing IOs have expanded into the field of education. Thus, more and more IOs can hold states accountable for their education policies today. The growth of the field of education IOs reached its peak around 20 years ago, and today there are around 30 organisations active. Moreover, the analysis also showed that education IOs use an entire spectrum of mechanisms to seek influence on their member states and beyond. These IOs have limited coercive means at their disposal; almost all are active in standard setting and some collect and provide comparative data. Therefore, IOs are an important element in the trajectory and translation of scripts in education.

First, we need to be aware of the number and character of IOs dealing with education policy. Although evidence indicates that the usual suspects – namely UNESCO, the World Bank and the OECD – are the most influential IOs in education and, therefore, deserve the broadest attention in academic studies due to their long-standing intrinsic mandate, financial opportunities and specific developed expertise, we must also recognise that there is still much we do not know about other actors in the field and about the extent to which they shape education policy in regions in which they are active. Only recently have additional case studies emerged on other actors, such as SEAMEO or ICESCO (e.g., Krogmann, 2022a, 2022b). The extent to which such actors focus on regional rather than global scripts in education requires further examination. By demanding specific characteristics for membership (e.g., cultural or religious value), such regional or transregional IOs persist through unique selling points in their respective bubbles. Following the theoretical approach of intrinsic features, regional or cultural education IOs thus find their niches in specific geographical or idiosyncratic contexts.

Second, IOs contribute data for accountability in education in a variety of ways. Although this study clearly shows the omnipresence of the OECD's PISA study as the global tool for generating and evaluating data, IOs not only design, produce and provide standardised tests across states, schools and at all levels of policy, they also provide other data, such as standards and ideas that influence policy-making across states.

The extent to which these standard-setting mechanisms impact future policies requires additional examination despite potential difficulties arising in the attribution of their causal effects. Such analyses would provide evidence of how much IOs can hold states accountable. Therefore, this study can only serve as the first step to research in this field. More systematic research needs to be conducted in general, particularly on less studied IOs and contextual factors. Do IOs consider regional factors such as religion, language, culture and traditional norms and values, and do they translate global scripts accordingly when they connect the global level with the local level of policy development?

As to theoretical insights, education studies, political science, sociology and international relations have all produced theoretical approaches that allow for an application to the field of education IO. These fields provide insights into the spread of IO influence, how their design impacts their opportunities for activities and the mechanisms by which they provide incentives for national policies. The empirical analysis in this study was guided by this framework, which allowed for a measurable determination of which IOs are involved in education and how they are involved in the education sector, revealing the characteristics of the world of education that IOs have introduced, such as the point of saturation in the growth of education IOs in the early 2000s with according "niches"; a globally even spread of education IOs across continents resembling a world culture; a majority of IOs having an institutional design of regionalist and economistic focus; little coercive power being exercised by education IOs (and if so, by global IOs and those with comprehensive responsibilities, such as the EU); and the dominance of PISA as an instrument of education evaluation used across the globe.

This chapter identifies some patterns, figures and facts about the field of education IOs and what IOs do. The extent to which IOs are challenged as global governance actors in today's world of uncertainty needs further examination (Jordana et al., 2024). This chapter's study is also limited because of its focus on intergovernmental actors; thus, it leaves out other important actors that require consideration and further examination. Expanding research to nongovernmental education IOs may indeed change or confirm the current picture of the literature. The Gates Foundation, the Soros Foundation or the Global Fund come to mind as the "usual suspects" in the realm of non-state actors. Many of these organisations have larger funds available than intergovernmental organisations and therefore possess significant tools for coercive mechanisms involving financial means. Overall, this contribution provides

more knowledge about the trajectory of education IOs. It shows that comparable work across policy fields is necessary to determine if and to what extent the population of education IOs resembles the activities of IOs in other contexts and whether education as a policy field has certain characteristics that differentiate it from other sectors.

References

Abbott, A. (1983). Sequences of social events: Concepts and methods for the analysis of order in social processes. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *16*(4), 129–147.

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, *21*, 93–113.

Abdul–Hamid, H. (2017). *Data for learning: Building a smart education data system*. The World Bank.

Abdul-Hamid, H., Saraogi, N., & Mintz, S. M. (2017). *Lessons learned from World Bank education management information system operations: Portfolio review, 1998-2014: SABER case study*. World Bank.

Adamson, F., Åstrand, B., & Darling-Hammond, L. (2016). *Global education reform*. Taylor & Francis.

Addey, C. & Sellar, S. (2018). Why do countries participate in PISA? Understanding the role of international large-scale assessments in global education policy. In A. Verger, M. Novelli, & H. K. Altinyelken (Eds.), *Global Education Policy and International Development* (2nd ed.) (pp. 97–117). Bloomsbury.

Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare: A Journal of Comparative and International Education*, *47*(3), 434–52.

Akkerman, S. F. & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research*, *81*(2), 132–169.

Allerup, P., Kovac, V., Kvåle, G., Langfeldt, G., & Skov, P. (2009). Evaluering av det nasjonale kvalitetsvurderingssystemet for grunnopplæringen [Evaluation of the national quality assessment system for basic education], Agderforskning, Kristiansand, Norway.

Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2010). Quality assurance, accountability, and qualification frameworks. In P. Altbach & L. Reisberg & L. Rumbley (Eds.),*Trends in Global Higher Education* (Vol. 22). Brill.

Ananny, M. & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.

Anderson, G. (2006). Assuring quality/resisting quality assurance: Academics' responses to 'quality' in some Australian universities. *Quality in Higher Education*, *12*(2), 161–173.

Arcia, G., Macdonald, K., Patrinos, H. A., & Porta, E. (2011). *School autonomy and accountability*. Human Development Education Network, The World Bank.

Armingeon, K. (2021). Intergovernmental organizations. In D. Béland, S. Leibfried, K.J. Morgan, H. Obinger, & C. Pierson (Eds.), *The Oxford Handbook of the Welfare* (2nd ed., pp. 417–432). Oxford University Press.

Augedal, P. K. (2006) *Målstyring og desentralisering–en øvelse i motsigelser?: vil det nasjonale kvalitetsvurderingssystemet sikre kvalitet i skolen innenfor dagens styringssystem?* [Management by objectives and decentralization – an exercise in contradictions?: Will the national quality assessment system ensure quality in schools within the current management system?]. Oslo, University of Oslo

Auld, E., Rappleye, J., & Morris, P. (2019). PISA for Development: How the OECD and World Bank shaped education governance post-2015. *Comparative Education*, *55*(2), 197–219.

Bacchi, C. (2009). *Analysing Policy*. Pearson Higher Education AU.

Bache, I., Bartle, I., Flinders, M., & Marsden, G. (2015). Blame games and climate change: Accountability, multi-level governance and carbon management. *British Journal of Politics and International Relations*, *17*, 64–88.

Baek, C. (2024). Navigating the terrain: global knowledge brokering and use in education policy. In C. Baek & G. Steiner-Khamsi (Eds.), *The Rise of Knowledge Brokers in Global Education Governance* (pp. 25–42). Edward Elgar Publishing.

Baek, C. & Steiner-Khamsi, G. (2024). Knowledge brokerage and global governance: Current research trends and future avenues for research. In C. Baek & G. Steiner-Khamsi (Eds.), *The Rise of Knowledge Brokers in Global Education Governance*, (pp. 190–198). NORRAG Series on International Education and Development.

Ball, S. J. (2018). The tragedy of state education in England: Reluctance, compromise and muddle – a system in disarray. *Journal of the British Academy*, *6*, 207–238.

Barnett, M.N. & Finnemore, M. (2004). *Rules for the World: International Organizations in Global Politics*. Cornell University Press.

Barry, A., Osborne, T., & Rose, N. (1996). Introduction. In A. Barry, T. Osborne, & N. Rose (Eds.), *Foucault and Political Reason* (pp. 1–17). Chicago University Press.

Bascia, N. & Osmond, P. (2013). Teacher union governmental relations in the context of educational reform. Education International.

Bellei, C. & Cabalin, C. (2013). Chilean student movement: Sustained struggle to transform a market-oriented educational system. *Current Issues in Comparative Education*, *15*(2), 108–123.

Bellei, C. & Vanni, X. (2015). Evolución de las políticas educacionales en Chile: 1980-2014. In C. Bellei (Ed.), *El gran experimento* (pp. 23–45). LOM.

Benavot, A. & Köseleci, N. (2015). Seeking quality in education: The growth of national learning assessments. *Paper Commissioned for the EFA Global Monitoring Report 2015.*

Benish, A. & Mattei, P. (2020). Accountability and hybridity in welfare governance. *Public Administration*, *98*, 281–290.

Benveniste, L. (2002). The political structuration of assessment: Negotiating state power and legitimacy. *Comparative Education Review*, *46*(1), 89–118.

Besche-Truthe, F. & Seitzer, H. (2025). Testing for the money: An analysis on the interdependence of participation in international large-scale assessments and development aid networks. *Frontiers in Education, 10.*

Bexell, M. (2005). *Exploring responsibility: Public and private in human rights protection* [PhD dissertation]. Department of Political Science, Lund University.

Bezes, P. & Parrado, S. (2013). Trajectories of administrative reform: Institutions, timing and choices in France and Spain. *West European Politics*, *36*(1), 22–50.

Bivins, T. H. (2006). Responsibility and accountability. In Fitzpatrick, K., & Bronstein, C. (Eds.), *Ethics in Public Relations: Responsible Advocacy* (pp. 19–38). Sage.

Blair, T. & Schröder, G. (1998). *Europe: The Third Way/Die Neue Mitte* [Working Documents No. 2]. Friedrich Ebert Foundation.

Bogotch, I. E. (2014). Autonomy and accountability: The power/knowledge knot. *School Leadership & Management*, *34*(4), 317–323.

Boli, J. & Thomas, G.M. (Eds.) (1999). *Constructing World Culture: International Nongovernmental Organizations Since 1875.* Stanford University Press.

Bonal, X., Pagès, M., Verger, A., & Zancajo, A. (2023). Regional policy trajectories in the Spanish education system: Different uses of relative autonomy. *Education Policy Analysis Archives*, *31*(68), 1–26.

Borowiak, C. T. (2011). *Accountability and Democracy: The Pitfalls and Promise of Popular Control*. Oxford University Press.

Boswell, C. (2015). The double life of targets in public policy: Disciplining and signalling in UK asylum policy. *Public Administration*, *93*(2), 490–505.

Bovens, M. (1998). *The Quest for Responsibility: Accountability and Citizenship in Complex Organizations*. Cambridge University Press.

Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics*, 33, 946–967.

Bovens, M. A. P., Goodin, R. E., & Schillemans, T. (Eds.). (2014). *The Oxford Handbook Public Accountability.* Oxford Handbooks.

Bowie, N. (1982). *Business Ethics.* Prentice-Hall.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, *82*(5), 977–1008.

Bromley, P., Furuta, J., Kijma, R., Overbey, L., Choi, M., & Santos, H. (2023). Global determinants of education reform, 1960 to 2017. *Sociology of Education*, *96*(2), 149–167.

Bromley, P., & Meyer, J. W. (2016). *Hyper–Organization. Global Organizational Expansion.* Oxford University Press.

Bromley, P., Nachtigal, T. & Kijima, R. (2024). Data as the new panacea: Trends in global education reforms, 1970–2018. *Comparative Education*, *60*(3), 401–422.

Bromley, P. & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *Academy of Management Annals*, *6*(1), 483–530.

Brunsson, N., Gustafsson Nordin, I., & Tamm Hallström, K. (2022). Organization: How more organization produces less responsibility. *Organization Theory*, *3*(4), 1–20.

Busemeyer, M.R. & Trampusch, C. (2011). Review article: Comparative political science and the study of education. *British Journal of Political Science*, *41*(2), 413–43.

Béland, D., Campbell, A. L., & Weaver, R. K. (2022). *Policy Feedback: How Policies Shape Politics*. Cambridge University Press.

Béland, D. & Cox, R. H. (2016). Ideas as coalition magnets: Coalition building, policy entrepreneurs, and power relations. *Journal of European Public Policy*, *23*(3), 428–445.

Cairney, P. (2019). *Understanding Public Policy: Theories and Issues* (vol. 2). Bloomsbury Publishing.

Campbell, J. L. & Pedersen, O. K. (Eds.). (2001). *The Rise of Neoliberalism and Institutional Analysis*. Princeton University Press.

Camphuijsen, M. K. & Levatino, A. (2022). Schools in the media: Framing national standardized testing in the Norwegian press, 2004-2018. *Discourse: Studies in the Cultural Politics of Education*, *43*(4), 601–616.

Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*, 36(5), 624–642.

Capano, G. (2018). Reconceptualizing layering – From mode of institutional change to mode of institutional design: Types and outputs. *Public Administration*, *97*(3), 590–604.

Capano, G. & Lippi, A. (2017). How policy instruments are chosen: Patterns of decision makers' choices. *Policy Sciences*, *50*(2), 269–293.

Capano, G., Zito, A. R., Toth, F., & Rayner, J. (2022). *Trajectories of Governance. How States Shaped Policy Sectors in the Neoliberal Age*. Palgrave.

Casserras, T. (2017). Per a què serveixen les auditories pedagògiques? [What are pedagogical audits for?] *Perspectiva Escolar*, 396, 64–67.

Cejudo, G. M. & Trein, P. (2023). Pathways to policy integration: A subsystem approach. *Policy Sciences*, *56*(1), 9–27.

Choi, M. (2024). Human rights and human capital discourse in national education reforms, 1960–2018. *Comparative Education Review*, *68*(1), 15–40.

Christensen, J. & Hesstvedt, S. (2024). Comparing ministerial evidence cultures: A quantitative analysis. *Policy and Society*, *43*(4), 494–520.

Christensen, T. (2006). Smart policy. *The Oxford Handbook of Public Policy*, *56*, 448–468.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284.

Clayton, G., Bingham, A. J., & Ecks, G. B. (2019). Characteristics of the opt-out movement: Early evidence for Colorado. *Education Policy Analysis Archives*, *27*(33).

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836.

Cohen, J. E. (2019). Review of Zuboff's the age of surveillance capitalism. *Surveillance and Society*, *17*(1/2), 240–245.

Cox, C. & Lemaitre, M. J. (1999). Market and state principles of reform in Chilean education: Policies and results. In G. Perry and D. Liepziger (Eds.) *Chile: Recent Policy Lessons and Emerging Challenges*, (pp. 149–188). World Bank.

Crawford, J. R. (2001). Teacher autonomy and accountability in charter schools. *Education and Urban Society*, *33*(2), 186–200.

Czarniawska-Joerges, B. & Sevón, G. (Eds.) (1996). *Translating Organizational Change* (vol. 56). Walter de Gruyter.

Darling-Hammond, L. (1994). National standards and assessments: Will they improve education? *American Journal of Education*, *102*(4), 478–510.

Darling-Hammond, L. (2010). Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching. *Center for American Progress*.

Debre, M., & Dijkstra, H. (2021). Institutional Design for a Post-Liberal Order: Why Some International Organizations Live Longer than Others. *European Journal of International Relations*, *27*(1), 311–39.

Diaz Rios, C. & Urbano–Canal, N. (2023). The World Bank and education policy in Colombia: A comparative analysis of the effects of international organizations' learning on domestic policy. *Journal of Comparative Policy Analysis: Research and Practice*, *25*(1), 101–117.

Dill, D. (2010). Quality assurance in higher education: Practices and issues. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 377–383). Elsevier Publications.

DiMaggio, P.J. & Powell, W.W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–60.

Dixson, D. D. & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice*, *55*(2), 153–159.

Dobbin, F., Simmons, B., & Garrett, G. (2007). The global diffusion of public policies: Social construction, coercion, competition, or learning? *Annual Review of Sociology*, *33*(1), 449–472.

Domina, T., Penner, A., & Penner, E. (2017). Categorical inequality: Schools as sorting machines. *Annual Review of Sociology*, *43*(1), 311–330.

Dubnick, M. J. (2011). Move over Daniel: We need some "accountability space". *Administration and Society*, *43*(6), 704–716.

Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. (2006). *Digital era governance: IT corporations, the state, and e-government*. Oxford University Press.

Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. (2005). New Public Management is dead – long live digital–era governance. *Journal of Public Administration Research and Theory*, *16*(3), 467–494.

Ehren, M. C., Gustafsson, J. E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, *51*(3), 375–400.

Eilstrup-Sangiovanni, M. (2020). Death of international organizations. The organizational ecology of intergovernmental organizations, 1815–2015. *Review of International Organisations*,*15*, 339–70.

Elfert, M. & Ydesen, C. (2023). *Global Governance of Education: The Historical and Contemporary Entanglements of UNESCO, the OECD and the World Bank*. Springer.

El-Khawas, E. (2007). Accountability and quality assurance: New issues for academic inquiry. In *International handbook of higher education* (pp. 23–37). Springer.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor.* New York: St. Martin's Press.

European Commission (2021). *Digital education action plan 2021-2027. Resetting Education, and training for the digital age*. https://education.ec.europa.eu/focus-topics/digital-education/plan

Falleti, T. G. (2010). *Decentralization and Subnational Politics in Latin America*. Cambridge University Press.

Finnemore, M. (1993). International organizations as teachers of norms: The United Nations Educational, Scientific, and Cultural Organization and science policy. *International Organization*, *47*(4), 565–597.

Finnemore, M. & Sikkink, K. (1998). International norm dynamics and political change. *International Organization*, *52*(4), 887–917.

Flinders, M. (2011). Daring to be a Daniel: The pathology of politicized accountability in a monitory democracy. *Administration and Society*, *43*(5), 595–619.

Flórez, M. T. P. (2023). Retrospective of 30 years of high-stakes assessment in Chile: The need for a paradigm shift. *Jornal de Políticas Educacionais*, *17*(2), 1–24.

Fontdevila, C. (2019). *Researching the adoption of school autonomy with accountability reforms: A methodological note on country case studies.* ReformEd Methodological Notes n1. doi:10.5281/zenodo.3557302.

Fontdevila, C., Verger, A., & Lopez-Leavy, M. (2024). *What's in a test: Tracing the evolution of national large-scale assessments systems.* [Seminar paper]. Developments and Changes in Education Systems across Global 'Cultural Spheres, Bremen.

Foucault, M. (1991). Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault Effect: Studies in Governmentality* (pp. 87–104). Harvester Wheatsheaf.

Fox, J. (2000). *Nonparametric Simple Regression: Smoothing Scatterplots* (issue 130). Sage.

Fox, N. J. (2011). Boundary objects, social meanings and the success of new technologies. *Sociology*, *45*(1), 70–85.

Fukuyama, F. (1992). *The End of History and the Last Man*. Penguin Books.

Furuta, J. (2022). The rationalization of 'education for all': The worldwide rise of national assessments, 1960–2011. *Comparative Education Review*, *66*(2), 228–252.

Furuta, J., Meyer, J. W., & Bromley, P. (2023). Education in a postliberal world society. In P. Mattei, X. Dumay, E. Mangez, & J. Behrend (Eds.), *The Oxford Handbook of Education and Globalization* (pp. 96–118). Oxford University Press.

Garritzmann, J. L. & K. Siderius. (2024). Introducing 'ministerial politics': Analyzingthe role and crucial redistributive impact of individual ministries in policy-making. *Governance*, *38*(1).

Gewirtz, S., Dickson, M., & Power, S. (2004). Unravelling a 'spun' policy: A case study of the constitutive role of 'spin' in the education policy process. *Journal of Education Policy, 19*(3), 321–342.

Glatter, R. (2012). Persistent preoccupations: The rise and rise of school autonomy and accountability in England. *Educational Management Administration and Leadership*, *40*(5), 559–575.

Goldfinch, S. F. (2023). The challenge of public administration reform: Introduction. In S. F. Goldfinch (Ed.), *Handbook of Public Administration Reform* (pp. 1–26). Edward Elgar Publishing.

Gray, J. (2018). Life, death, or zombie? The vitality of international organizations. *International Studies Quarterly*, *62*(1), 1–13.

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy*, *24*(1), 23–37.

Grek, S. (2014). OECD as a site of coproduction: European education governance and the new politics of 'policy mobilization'. *Critical Policy Studies*, *8*(3), 266–81.

Gunter, H. M., Hall, D., & Mills, C. (2014). Consultants, consultancy and consultocracy in education policymaking in England. *Journal of Education Policy*, *30*(4), 518–539.

Guskey, T. R. (2007). Using assessments to improve teaching and learning. Ahead of the Curve. In D. Reeves (Eds.), *The Power of Assessment to Transform Teaching and Learning*, (pp. 15–29). Solution Tree Press.

Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science*, *3*(10), 833–838.

Hall, N. (2016). *Displacement, Development, and Climate Change: International Organizations Moving Beyond Their Mandates*. Routledge.

Hall, P. A. 1993. "Policy paradigms, social learning, and the state: the case of economic policymaking in Britain." Comparative Politics, 25(3): 275-296.

Hargreaves, A., Braun, H., & Gebhardt, K. (2013). Data-driven improvement and accountability. National Education Policy Center, University of Colorado.

Harlow C., & Rawlings R. (2007). Promoting accountability in multi-level governance: A network approach. *European Law Journal*, *13*, 542–562.

Hart H. L. A. (2008). *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed.). Oxford University Press.

Hartong, S. & Nikolai, R. (2017). Observing the "local globalness" of policy transfer in education. *Comparative Education Review*, *61*(3), 519–537.

Hartong, S., Decuypere, M., & Lewis, S. (2024). Disentangling the temporalities of digital and predictive governance: Rhythmanalysis as a methodological framework. *Time and Society*, *34*(2), 178–201.

Hawkins, D.G., Lake, D.A., Nielson, D., & Tierney, M.J. (2006). Delegation under anarchy: The states, international organizations, and principal-agent theory. In D. Hawkins, D. A. Lake, D. Nielson, & M. J. Tierney (Eds.), *Delegation and Agency in International Organizations* (pp. 3–38). Cambridge University Press.

Heimstädt M. & Dobusch L. (2020). Transparency and accountability: Causal, critical and constructive perspectives. *Organization Theory*, *1*, 1–12.

Henri, M., Johnson, M. D., & Nepal, B. (2017). A review of competency-based learning: Tools, assessments, and recommendations. *Journal of Engineering Education, 106*(4), 607–638.

Högberg, B. & Lindgren, J. (2021). Outcome-based accountability regimes in OECD countries: A global policy model? *Comparative Education*, *57*(3), 301–321.

Holbein, J. B. & Ladd, H. F. (2017). Accountability pressure: Regression discontinuity estimates of how No Child Left Behind influenced student behavior. *Economics of Education Review, 58*, 55–67.

Hollyer, J. R., Rosendorff, P., & Vreeland, J. R. (2011). Democracy and transparency. *The Journal of Politics*, *73*(4), 1191–1205.

Honey, N. & Carrasco, A. (2023). A new admission system in Chile and its foreseen moderate impact on access for low-income students. *Educational Evaluation and Policy Analysis, 45*(1), 108–133.

Hood, C. (1983). *Tools of Government*. Macmillan.

Hood, C. (2007). Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance*, *20*(1), 127–144.

Hooghe, L., Lenz, T., & Marks, G. (2019). *A Theory of International Organization*. Oxford University Press.

Hossain, M. (2022). Diffusing "destandardization" reforms across educational systems in low- and middle-income countries: The case of the World Bank, 1965 to 2020. *Sociology of Education*, *95*(4), 320–339.

Hossain, M. (2023). Large–scale data gathering: Exploring World Bank's influence on national learning assessments in LMICs. *International Journal of Educational Development*, *102*, 1–8.

Hovdenak, S. S. & Stray, J. H. (2015). *Hva skjer med skolen? En kunnskapssosiologisk analyse av norsk utdanningspolitikk fra 1990-tallet og frem til i dag [What's happening to school? A sociology of knowledge analysis of Norwegian educational policy from the 1990s to the present day].* Fagbokforlaget.

Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal*, *31*(5), 605–622.

Inspecció d'Educació, (2022). *Pla Director de la Inspecció d'Educació 2021-2025 [Education Inspectorate Master Plan 2021-2025]*. Departament d'Educació.

Jabbar, H., Chanin, J., Haynes, J., & Slaughter, S. (2020). Teacher power and the politics of union organizing in the charter sector. *Educational Policy*, *34*(1), 211–238.

Jacobson, H. K. (1979). *Networks of Interdependence: International Organizations and the Global Political System*. Alfred A. Knopf.

Jakobi, A. P., Martens, K., & Wolf, K. D. (Eds). (2010). *Education in Political Science: Discovering a Neglected Field*. Routledge.

Jarke, J. & Breiter, A. (2019). Editorial: The datafication of education. *Learning, Media and Technology*, *44*(1), 1–6.

Jones, P. W. (1988). *International Policies for Third World Education: UNESCO, Literacy and Development*. Routledge.

Jones, P. W. (2007). *World Bank Financing of Education: Lending, Learning and Development (2nd ed)*. Routledge.

Jordana, J., Holesch, A., et al. (2024). Institutional transformations of global governance – key challenges for international organisations. *International Review of Public Policy*, *6*(3).

Jungblut, J., Maltais, M., Ness, E., & Rexe, D. (Eds). (2023). *Comparative Higher Education Politics: Policymaking in North America and Western Europe*. Springer.

Kabanov, Y. & Karyagin, M. (2018). Data-driven authoritarianism: Non-democracies and big data. In *Digital Transformation and Global Society: Third International Conference, DTGS 2018, St. Petersburg, Russia, May 30–June 2, 2018, Revised Selected Papers, Part I 3* (pp. 144–155). Springer International Publishing.

Kamens, D. H. & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, *54*(1), 5–25.

Kassim, H. & Le Galès, P. (2010). Exploring governance in a multi-level polity: A policy instruments approach. *West European Politics*, *33*(1), 1–21.

Keck, M. E. & Sikkink, K. (1998). *Activists Beyond Borders: Advocacy Networks in International Politics* (vol. 35). Cambridge University Press.

Keddie, A. (2015). School autonomy, accountability and collaboration: A critical review. *Journal of Educational Administration and History*, *47*(1), 1–17.

Kijima, R. & Leer, J. (2015). Legitimacy, state-building, and contestation in education policy development: Chile's involvement in cross-national assessments. In W. Smith (Eds.), *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice* (pp. 43-62). Oxford Studies in Comparative Education.

Kijima, R. & Lipscy, P. Y. (2020). International assessments and education policy: Evidence from an elite survey. In J. Kelley & B. Simmons (Eds.), *The Power of Global Performance Indicators* (pp. 174–202). Cambridge University Press.

Kijima, R. & Lipscy, P. Y. (2023). Competition and regime complex architecture: Authority relations and differentiation in international education. *Review of International Political Economy*, *30*(6), 2150–2177.

Kijima, R. & Lipscy, P. Y. (2024). The politics of international testing. *The Review of International Organizations*, *19*(1), 1–31.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*(3), 255–276.

Koremenos, B., Lipson, C., & Snidal, D. (2001). The rational design of international institutions. *International Organization*, *55*(4), 761–99.

Krill de Capello, H. H. (1970). The creation of UNESCO. *International Organization*, *24*(1), 1–30.

Krogh, A. H. & Triantafillou, P. (2024). Developing New Public Governance as a public management reform model. *Public Management Review*, *26*(10), 3040–3056.

Krogmann, D. (2022a). Regional ideas in international education organizations: The case of SEAMEO. In K. Martens & M. Windzio (Eds.), *Global Pathways to Education. Cultural Spheres, Networks, and International Organizations* (pp. 217–237). Palgrave Macmillan.

Lake, D. A., Martin, L. L., & Risse, T. (2021). Challenges to the liberal order: Reflections on international organization. *International Organization*, *75*(2), 225–257.

Langfeldt, G. & Lauvdal, T. (2006). *Evaluering av det nasjonale kvalitetsvurderingssystemet [Evaluation of the national quality assessment system],* Interim report 1.4.

Lascoumes, P. & Le Galès, P. (2007). Introduction: Understanding public policy through its instruments—from the nature of instruments to the sociology of public policy instrumentation. *Governance*, *20*(1), 1–21.

Le Galès, P. (2016). Performance Measurement as a policy instrument. *Policy Studies*, *37*(6), 508–520.

Lee, C. K., & Strang, D. (2006). The international diffusion of public-sector downsizing: Network emulation and theory-driven learning. *International Organization*, *60*(4), 883–909.

Lenz, T., Bezuijen, J., Hooghe, L., & Marks, G. (2015). Patterns of international organization: Task specific vs. general purpose. *Politische Vierteljahresschrift, Special Issue 49*, 131–56.

Leonelli, S. (2016). Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160122.

Lerch, J.C. & Buckner, E. (2018). From education for peace to education in conflict: Changes in UNESCO discourse, 1945–2015. *Globalisation, Societies andEducation*, *16*(1), 27–48.

Levatino, A., Verger, A., Camphuijsen, N., Termes, A., & Parcerisa, L. (2024). School governance through performance-based accountability: A comparative analysis of its side effects across different regulatory regimes. *Review of Research in Education*, *48*, 248–286.

Lindquist, L. & Llewellyn, S. (2003). Accountability, responsibility and organization. *Scandinavian Journal of Management*, *19*, 251–273.

Lindseth, A. A. & Hanushek, E. A. (2009). *Schoolhouses, Courthouses, and Statehouses: Solving the Funding-Achievement Puzzle in America's Public Schools*. Princeton University Press.

Lo, W. Y. W. (2014). Think global, think local: The changing landscape of higher education and the role of quality assurance in Singapore. *Policy and Society*, *33*(3), 263–273.

Lockheed, M. (2015). *Why do countries participate in international large-scale assessments? The case of PISA* [World Bank policy research working paper, 7447, October 19, 2015]. The World Bank.

Loureiro, A. & Cruz, L. (2020). *Achieving World-Class Education in Adverse Socioeconomic Conditions* [World Bank policy paper]. The World Bank.

Lubienski, C. (2019). Advocacy networks and market models for education. In M. Parreira do Amaral, G. Steiner-Khamsi & C. Thompson (Eds.), *Researching the Global Education Industry: Commodification, the Market and Business Involvement* (pp. 69–86). Springer International Publishing.

Lubienski, C., Malin. J. R., & Rowe, E. (2024). Evidence production, promotion, and pollution: private interests and knowledge brokerage in US education policy. In C. Baek & G. Steiner-Khamsi (Eds.), *The Rise of Knowledge Brokers in Global Education Governance* (pp. 171–189). Edward Elgar Publishing.

Lubienski, C., Scott, J., & DeBray, E. (2014). The politics of research production, promotion, and utilization in educational policy. *Educational Policy*, *28*(2), 131–144.

Mahoney, J. & Thelen, K. (2010). A theory of gradual institutional change. *Explaining Institutional Change: Ambiguity, Agency, and Power*, *1*(1).

Mahoney, J. & Thelen, K. (Eds.). (2015). *Advances in Comparative-historical Analysis*. Cambridge University Press.

March, J. G. & Olsen, J. P. (1995). *Democratic Governance*. The Free Press.

Marianno, B. D., Woo, D. S., & Kennedy, K. (2023). Collective bargaining agreement restrictiveness in unionized charter schools. *Educational Policy*, *38*(5), 1044–1076.

Markiewicz, A. & Patrick, I. (2015). *Developing Monitoring and Evaluation Frameworks*. Sage Publications.

Maroy, C. (2012). Towards post-bureaucratic modes of governance: A European perspective. In G. Steiner-Khamsi & F. Waldow (Eds.), *World Yearbook of Education 2012: Policy Borrowing and Lending* (pp. 63–79). Routledge.

Maroy, C. & Pons, X. (2021). Re-conceptualising education policy trajectories in a globalised world: Lessons from a multi-level comparison of accountability in France and Quebec. *Comparative Education*, *57*(4), 560–578.

Martens, K. (2007). How to become an influential actor – the 'comparative turn' in OECD education policy. In K. Martens, A. Rusconi, & K. Leuze (Eds.), *New Arenas of Education Governance: The Impact of International Organizations and Markets on Education Policy Making* (pp. 40–56). Palgrave Macmillan.

Martens, K. & Jakobi, A.P. (2010). Introduction – the OECD as an actor in international politics. In K. Martens & A. P. Jakobi (Eds.), *Mechanisms of OECD Governance – International Incentives for National Policy Making?* (pp. 1–25). Oxford University Press.

Martens, K., Niemann, D. & Krogmann, D. (2024). The expansion of education in and across international organizations. In P. Mattei, E. Mangez, J. Behrend,

& X. Dumay (Eds.), *Handbook on Globalization and Education.* Oxford University Press.
Martens, K., Niemann, D., & Teltemann, J. (2016). Effects of international assessments in education–a multidisciplinary review. *European Educational Research Journal*, *15*(5), 516–522.
Mattei, P. (2023). *Democraticizing Science: The Political Roots of Public Engagement Agenda*. Bristol University Press.
Mausethagen, S., Prøitz, T. S., & Skedsmo, G. (2020). Redefining public values: Data use and value dilemmas in education. *Education Inquiry*, *12*(1), 1–16.
Mazzucato, M. & Collington, R. (2023). *The Big Con: How the Consulting Industry Weakens our Business, Infantilizes our Governments and Warps our Economies.* Allen Lane.
McGimpsey, I. (2017). Late neoliberalism: Delineating a policy regime. *Critical Social Policy*, *37*(1), 64–84.
McNeely, C. L. (1995). Prescribing national education policies: The role of international organizations. *Comparative Education Review*, *39*(4), 483–507.
McNeely, C. L. & Cha, Y.-K. (1994). Worldwide educational convergence through international organisations: Avenues for research. *Education Policy Analysis Archives*, *2*(14).
Meckes, L. (2003). El Simce: Su desarrollo y desafíos actuales. *Pensamiento Educativo*, *33*(2), 160–178.
Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, *30*(2), 521–546.
Merry, S. (2009). Relating to the subjects of human rights: The culture of agency in human rights discourse. In M. Freeman & D. Napier (Eds.), *Law and Anthropology* (pp. 385–407). Oxford University Press.
Messner, M. (2009). The limits of accountability. *Accounting, Organizations and Society*, *34*, 918–938.
Meyer, J. W. & Bromley, P. (2013). The worldwide expansion of 'organization'. *Sociological Theory*, *31*(4), 366–89.
Meyer, J. W. & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *Journal of Economic Sociology*, *12*(1), 43–67.
Meyer, J. W., Boli, J., Thomas, G. M., & Ramirez, F. O. (1997). World society and the nation state. *American Journal of Sociology*, *103*(1), 144–81.
Miller D. (2001). Distributing responsibilities. *Journal of Political Philosophy*, *9*, 453–471.
Ministry of Finance (1995). *Agreement between the Ministry of Health and Social Security and the Radiation Protection Agency, the Ministry of Industry and the Building Research Institute, the Ministry of Education and the Women's School in Reykjavík, the Ministry of Transport and Institute of Lighthouses and Ports, the Ministry of Agriculture and the Farm school in Hólar.* Ministry of Finance of Iceland.
Mizala, A. & Schneider, B. (2020). Promoting quality education in Chile: The politics of reforming teacher careers. *Journal of Education Policy*, *35*(4), 529–555.

Morley, L. (2003). *Quality and power in higher education*. Society for Research into Higher Education and Open University Press.

Moutsios, S. (2009). International organisations and transnational education policy. *Compare: A Journal of Comparative and International Education*, *39*(4), 469–481.

Mundy, K. (1999). Educational multilateralism and world (dis)order. *Comparative Education Review*, *42*(4), 448–478.

Mundy, K. (2007). Educational multilateralism – Origins and indications for global governance. In K. Martens, A. Rusconi, & K. Leuze (Eds.), *New Arenas of Education Governance. The Impact of International Organizations and Markets on Educational Policy Making* (pp. 19–39). Palgrave Macmillan.

Mundy, K. & Menashy, F. (2014). The World Bank and private provision of schooling: A look through the lens of sociological theories of organizational hypocrisy. *Comparative Education Review*, *58*(3), 401–27.

Mundy, K. & Verger, A. (2015). The World Bank and the global governance of education in a changing world order. *International Journal of Educational Development*, *40*, 9–18.

Nagel, A, Martens, K., & Windzio, M. (2010). Introduction – education policy in transformation. In K. Martens, A. Nagel, M. Windzio, & A. Weymann (Eds.), *Transformation of Education Policy* (pp. 3–27). Palgrave Macmillan.

Nielson, D. L. & Tierney, M. J. (2003). Delegation to international organizations: Agency theory and World Bank reform. *International Organization*, *57*(2), 241–76.

Niemann, D. (2012). *The soft, the hard, and the complementary? The impact of international organizations on national policy making* [Paper presentation at the International Studies Association Meeting].

Niemann, D. (2021). International organizations in education: New takes on old paradigms. In M. Windzio & K. Martens (Eds.), *The Global Development, Diffusion, and Transformation of Education Systems* (pp. 127–61). Palgrave Macmillan.

Niemann, D., Krogmann, D., & Martens, K. (2023). Torn into the abyss? Subpopulations of international organizations in times of a declining liberal international order. *Global Governance: A Review of Multilateralism and International Organizations*, *29*(3), 271–94.

Niemann, D. & Martens, K. (2018). Soft governance by hard facts? The OECD as a knowledge broker in global education policy. *Global Social Policy*, *18*(3), 267–83.

Niemann, D. & Martens, K. (2021). Global discourses, regional framings and individual showcasing: Analyzing the world of education IOs. In K. Martens, D. Niemann, & A. Kaasch (Eds.), *International Organizations in Global Social Governance* (pp. 163–86). Palgrave Macmillan.

Niemann, D., Martens, K., & Teltemann, J. (2017). PISA and its consequences: Shaping education policies through international comparisons. *European Journal of Education*, *52*(2), 175–183.

Norwegian Directorate for Education and Training (NDET). (2011). *OECD review on evaluation and assessment frameworks for improving school outcomes: Country background report for Norway.* NDET website: https://

www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2011/5/oecd_country_report_norway.pdf
Nóvoa, A. & Yariv-Mashal, T. (2003). Comparative research in education: A model of governance or a historical journey? *Comparative Education*, *39*(4), 423–438.
Nusche, D., Earl, L., Maxwell, W., & Shewbridge, C. (2011). *OECD reviews of evaluation and assessment in education: Norway 2011.* OECD Publishing.
Nye Jr, J. S. (2017). Will the liberal order survive: The history of an idea. *Foreign Affairs*, *96*, 10.
OECD. (2019). *OECD future of education and skills 2030. OECD learning compass 2030. A series of concept notes*. OECD.
OECD. (2020). *Education Policy Outlook: Turkey.* OECD. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-policy-outlook/country-profile-Turkey-2020.pdf
OECD. (2021). *OECD survey on social and emotional skills technical report*. OECD.
OECD. (2022). *The OECD good practice principles for public service design and delivery in the digital age* [OECD Public Governance Policy Papers, No. 23]. OECD.
OECD. (2025). *Education Policy Outlook*. OECD.
Olsen, J. P. (2009). Democratic government, institutional autonomy and the dynamics of change. *West European Politics*, *32*(3), 439–465.
Olsen, J. P. (2015). Democratic order, autonomy, and accountability. *Governance: An International Journal of Policy, Administration, and Institutions*, *28*(4), 425–440.
Osborne, S. P. (2010). *The New Public Governance? Emerging Perspectives on the Theory and Practice of Public Governance*. Routledge.
Ostry, J. D., Loungani, P., & Furceri, D. (2016). Neoliberalism oversold? *Finance and Development*, *53*(2), 38–41.
Parcerisa, L. & Falabella, A. (2017). The consolidation of the evaluative state through accountability policies: Trajectory, enactment and tensions in the Chilean education system. *Education Policy Analysis Archives*, *25*(89).
Parcerisa, L., Pagès, M., Termes, A., & Collet-Sabé, J. (2022). Why do opt-out movements succeed (or fail) in low-stakes accountability systems? A case study of the Network of Dissident Schools in Catalonia. *Education Policy Analysis Archives*, *30*, (133).
Parcerisa, L., Verger, A., & Browes, N. (2022). Teacher autonomy in the age of performance-based accountability: A review based on teaching profession regulatory models (2017-2020). *Education Policy Analysis Archives*, *30*(100).
Parcerisa, L., Verger, A., & Falabella, A. (2020). High-stakes accountability and the expansion of a school improvement industry in Chile: A public-private sector comparison. In A. Hogan & G. Thompson (Eds.), *Privatisation and Commercialisation in Public Education* (pp. 119–133). Routledge.
Peters, B. G. (2002). *The politics of tool choice. The tools of government: A guide to the new governance.* In Salamon, L. M. (Ed.). *The Tools of Government: A Guide to the New Governance* (pp. 552–564). Oxford University Press.

Peters, B. G. & Pierre, J. (2016). *Comparative Governance: Rediscovering the Functional Dimension of Governing*. Cambridge University Press.
Pierson, P. (2004). *Politics in Time: History, Institutions, and Social Analysis*. Princeton University Press.
Pierson, P. (2015). Power and path dependence. In J. Mahoney & K. Thelen (Eds.), *Advances in Comparative–historical Analysis. Strategies for Social Inquiry* (pp. 123–146). Cambridge University Press.
Pizmony-Levy, O. (2013). *Testing for All: The Emergence and Development of International Assessment of Student Achievement, 1958–2012*. Indiana University ProQuest Dissertations Publishing.
Podgursky, M. J. & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, *26*(4), 909–949.
Pollitt, C., & Bouckaert, G. (2011). *Public Management Reform: A Comparative Analysis – New Public Management, Governance, and the Neo-weberian State* (3rd ed.). Oxford University Press.
Porto de Oliveira, O. (2017). *International Policy Diffusion and Participatory Budgeting: Ambassadors of Participation, International Institutions And Transnational Networks*. Palgrave.
Power, M. (1997). *The Audit Society: Rituals of Verification*. OUP Oxford.
Power, M. (2022). Theorizing the economy of traces: From audit society to surveillance capitalism. *Organization Theory*, *3*(3).
Quilabert, E., Verger, A., Moschetti, M. C., Ferrer-Esteban, G., & Pagès, M. (2024). The obstacle race to educational improvement: Governance, policies, and practices in disadvantaged schools. *British Journal of Sociology of Education*, *45*(6), 934–956.
Ramirez, F. O., Schofer, E., & Meyer, J. W. (2018). International tests, national assessments, and educational development (1970–2012). *Comparative Education Review*, *62*(3), 344–364.
Resnik, J. (2006). International organizations, the 'education-economic growth' black box, and the development of world education culture. *Comparative Education Review*, *50*(2), 173–195.
Rittberger, V. (1995). *Anpassung oder Austritt: Industriestaaten in der UNESCO-Krise. Ein Beitrag zur vergleichenden Außenpolitikforschung*. Ed. Sigma.
Rittberger, V., Zangl, B., & Kruck, A. (2019). *International Organization* (3rd ed). Palgrave Macmillan.
Rizvi, F. & Lingard, B. (2009). *Globalizing Education Policy* (1st ed.). Routledge.
Roberts, J. (1991). The possibilities of accountability. *Accounting, Organizations and Society*, *16*, 355–368.
Roberts J. & Scapens, R. (1985). Accounting systems and systems of accounting: Understanding accounting practices in their organizational contexts. *Accounting, Organizations and Society*, *10*, 443–456.
Robertson, S. L., Bonal, X., & Dale, R. (2002). GATS and the education service industry: The politics of scale and global reterritorialization. *Comparative Education Review*, *46*(4), 472–95.

Rose, N. (1992). Governing the enterprising self. In P. Heelas & P. Morris (Eds.), *The Values of the Enterprise Culture: The Moral Debate* (pp. 141–164). Routledge.

Rose, N. (1999). *Powers of Freedom. Reframing Political Thought*. Cambridge University Press.

Rose, N., O'Malley, P., & Valverde, M. (2006). Governmentality. *Annual Review of Law and Social Science*, *2*, 83–104.

Rosser, A., King, P., & Widoyoko, D. (2022). *The Political Economy of the Learning Crisis in Indonesia. Research on improving systems of education*. RISE Political Economy Paper 01.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data and Society*, *6*(1).

Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change*, *7*(4), 259–287.

Sahlin, K., & Wedlin, L. (2008). Circulating ideas: imitation, translation and editing. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin (Eds.), *Circulating Ideas: Imitation, Translation and Editing* (pp. 218–242). SAGE Publications.

Salganik, L. H., Rychen, D. S., Moser, U., & Konstant J. W. (1999). *Definition and Selection Of Competencies. Projects on Competencies in the OECD Context. Analysis of Theoretical and Conceptual Foundations*. Swiss Federal Statistics Office.

Salter, B. & Tapper, T. (2000). The Politics of Governance in Higher Education: The Case of Quality Assurance. *Political studies*, *48*(1), 66–87.

Santibañez, L. (2006). Why we should care if teachers get A's: Teacher test scores and student achievement in Mexico. *Economics of Education Review*, *25*(5), 510–520.

Sanyal, B. C., & Martin, M. (2007). Quality assurance and the role of accreditation: An overview. In Global University Network for Innovation Report (Ed.), *Higher Education in the World 2007: Accreditation for Quality Assurance: What is at Stake?* (pp. 3–17). Palgrave Macmillan.

Schmidt, V. E. (2009). Putting the political back into political economy by bringing the state back in yet again. *World Politics*, *61*(3), 516–546.

Seawright, J. & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, *61*(2), 294–308.

Segura, J., Gairín, J., & Silva, P. (2021). Implicaciones de la Inspección Educativa en Cataluña en el Proceso de Autonomía de Centros [Implications of educational inspection in Catalonia in the process of school autonomy]. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, *19*(1), 61–82.

Seitzer, H., Baek, C., & Steiner-Khamsi, G. (2023). Instruments of lesson-drawing: Comparing the knowledge brokerage of the OECD and the World Bank. *Policy Studies*, *45*(6), 839–859

Sellar, S. & Lingard, B. (2014). The OECD and the expansion of PISA: new global modes of governance in education. *British Educational Research Journal*, *40*(6), 917–936.

Shephard, D. & Delprato, M. (2024). Introduction to missing education data and the SDG 4 data regime. In M. Delprato & D. Shephard (Eds.), *Achieving equitable education. Missing education data and the SDG4 data regime* (pp. 1–13). Edward Elgar Publishing.

Shivakoti, R. & Howlett, M. (2022). Agenda–setting tools in theory and practice. In M. Howlett (Ed.), *Routledge Handbook of Policy Tools*, (pp. 113–124). Routledge.

Shore, C. & Wright, S. (2011). Conceptualising policy: Technologies of governance and the politics of visibility. In C. Shore, S. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the Analysis of Contemporary Powe*r (pp. 1–26). Berghahn Books.

Simeonova, R., Parvanova, Y., Brown, M., McNamara, G., Gardezi, S., Hara, J. O., & Blanco, C. (2020). A continuum of approaches to school inspections: Cases from Europe. *Pedagogy*, *92*(4), 487–506.

Simons, A. & Voß, J. P. (2018). The concept of instrument constituencies: Accounting for dynamics and practices of knowing governance. *Policy and Society*, *37*(1), 14–35.

Star, S. L. & Griesemer, J. R. (1989). Institutional ecology, translations, and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, *19*(3), 387–420.

Stark, D. & Broeck, P. V. (2024). Principles of algorithmic management. *Organization Theory*, *5*(2).

Stark, D. & Pais, I. (2020). Algorithmic management in the platform economy. *Sociologica*, *14*(3), 47–72.

Steinebach, Y., Knill, C., & Severin, C. (2024). International organizations and policy implementation. In F. Sager, C. Mavrot, & L. R. Keiser (Eds.), *Handbook of Public Policy Implementation* (pp. 407–17). Edward Elgar Publishing.

Steiner-Khamsi, G. (2012). Understanding policy borrowing and lending: Building comparative policy studies. In G. Steiner-Khamsi and F. Waldow (Eds.), *World Yearbook of Education 2012: Policy Borrowing and Lending in Education* (1st ed., pp. 3–17). Routledge.

Steiner-Khamsi, G. (2016). New directions in policy borrowing research. *Asia Pacific Education Review*, *17*(3), 381–390.

Steiner-Khamsi, G. (2021a). Externalisation and structural coupling: Applications in comparative policy studies in education. *European Educational Research Journal*, *20*(6), 806–820.

Steiner-Khamsi, G. (2025). *Time in Education Policy Transfer. The Seven Temporalities of Global School Reform.* Palgrave.

Steiner-Khamsi, G., Faul, M. V., Baek, C., Iwabuchi, K., & Hopkins, A. N. (2024). *Strategic review of global and regional evidence and knowledge initiatives, networks and platforms in education.* UNESCO.

Steiner-Khamsi, G., Jóhannesdóttir, K., Magnúsdóttir, B. R. (2024). The school-autonomy-with-accountability reform in Iceland: Looking back and making sense. *Nordic Journal of Studies in Educational Policy*,1–19.

Steiner-Khamsi, G., Karseth, B., & Baek, C. (2020). From science to politics: Commissioned reports and their political translation into white papers. *Journal of Education Policy*, *35*(1), 119–144.

Steiner-Khamsi, G., Martens, K., & Ydesen, C. (2024). Governance by numbers 2.0: Policy brokerage as an instrument of global governance in the era of information overload. *Comparative Education*, *60*(4), 537–554.

Steiner-Khamsi, G., & Waldow, F. (2018). PISA for scandalisation, PISA for projection: The use of international large-scale assessments in education policy making–an introduction. *Globalisation, Societies and Education*, *16*(5), 557–565.

Stone, D. (2012). Transfer and translation of policy. *Policy Studies*, *33*(6), 483–499.

Supovitz, J. (2009). Can high-stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, *10*, 211–227.

Thelen, K. (2004). *How Institutions Evolve*. Cambridge University Press.

Tilly, C. (1989). *Big Structures, Large Processes, Huge Comparisons*. Russell Sage.

Tomaskovic-Devey, D., Avent-Holt, D., Zimmer, C., & Harding, S. (2009). The categorical generation of organizational inequality: A comparative test of Tilly's durable inequality. *Research in Social Stratification and Mobility*, *27*(3), 128–142.

Trnka, S. & Trundle, C. (2014). Competing responsibilities: Moving beyond neoliberal responsibilisation. *Anthropological Forum*, *24*(2), 136–153.

UIS (UNESCO Institute for Statistics) (2024). *Background information on education statistics in the UIS database*. UIS.

UNESCO (2000). *Education for All: Meeting our Collective Commitments*. World Education Forum, Dakar, Senegal, 26–28 April 2000.

UNESCO (2022). *Global education monitoring report: South Asia: Non-state actors in education: Who chooses? Who loses?* UNESCO.

UNICEF (2023). *Background paper on the development of a socio-emotional skills assessment in the Philippines: Global and local agenda*. UNICEF.

Valenzuela, J. P. & Montecinos, C. (2017). Structural reforms and equity in Chilean schools. In *Oxford Research Encyclopedia of Education* (pp. 1–25). Oxford University Press.

Van Gunsteren, H. R. (1976). *The quest for control: A critique of the rational-central-rule approach in public affairs*. John Wiley.

Vera, S. (2011). Cronología del Conflicto: El Movimiento Estudiantil en Chile [The student movement in Chile], 2011. *Anuario del Conflicto Social*, *1*(1), 252–261.

Verger, A. (2009). *WTO/GATS and the Global Politics of Higher Education*. Routledge.

Verger, A., Ferrer-Estaban, G., Quilabert, E., Moschetti, M., & Levatino, A. (2024). *Escoles i professorat davant l'autonomia de centre: Reptes de l'avaluació, la millora i la governança escolar [Schools and teachers facing school autonomy: Challenges of evaluation, improvement and school governance]*. Fundació Bofill.

Verger, A., Fontdevila, C., & Parcerisa, L. (2019a). Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. *Discourse: Studies in the Cultural Politics of Education*, *40*(2), 248–270.

Verger, A., Fontdevila, C., & Parcerisa, L. (2019b). Constructing school autonomy with accountability as a global policy model: A focus on OECD's governance mechanisms. In C. Ydesen (Ed.), *The OECD's Historical Rise in Education: The Formation of a Global Governing Complex* (pp. 219–243). Springer International Publishing.

Verger, A., Parcerisa, L., & Fontdevila, C. (2019c). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, *71*(1), 5–30.

Verger, A., Fontdevila, C., & Zancajo, A. (2017). Multiple paths towards education privatization in a globalizing world: A cultural political economy review. *Journal of Education Policy*, *32*(6), 757–787.

Voisin, A. (2020). Siren song: Performance-based accountability systems, effectiveness, and equity: Evidence from PISA in education systems in Europe and Canada. In *World Yearbook of Education 2021* (pp. 215–240). Routledge.

Voisin, A. & Dumay, X. (2020). How do educational systems regulate the teaching profession and teachers' work? A typological approach to institutional foundations and models of regulation, *Teaching and Teacher Education*, *96*, 103–144.

Von der Embse, N. P., Pendergast, L. L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states, *Teaching and Teacher Education*, *59*, 492–502.

Vögtle, E. M. & Martens, K. (2014). The Bologna Process as a template for transnational policy coordination. *Policy Studies*, *35*(3), 246–63.

Waldow, F. (2009). What PISA did and did not do: Germany after the 'PISA-shock'. *European Educational Research Journal*, *8*(3), 476–483.

Weiner, L. (2013). Social justice teacher activism and social movement unionism: Tensions, synergies, and space, *Multidisciplinary Journal of Educational Research, 3*(3).

Wimmer, A. (2023). Layered legacies. How multiple histories shaped the attitudes of contemporary Europeans, *Sociological Science*, *10*, 1–46.

World Bank (2008). *Brazil. Toward a more inclusive and effective participatory budget in Porto Alegre. Report No. 4-144-BR*. World Bank.

World Bank (2021). *What is learning poverty? Brief, April 21.* World Bank.

World Bank (2022). *GovTech. Putting people first.* World Bank.

World Bank (2024). *World Bank EdStats Query.* World Bank.

World Bank (2025a). *Civic and citizen engagement*. World Bank.

World Bank (2025b). *Systems approach for better education results (SABER).* World Bank.

World Economic Forum (2016). *The Future of jobs employment, skills and workforce strategy for the fourth industrial revolution: Global challenge insight report*. WEF.

Yeung, K. (2023). The new public analytics as an emerging paradigm in public sector administration. *Tilburg Law Review, 27*(2), 1–32.
Zancajo, A., Verger, A., & Bolea, P. (2022). Digitalization and beyond: The effects of Covid-19 on post-pandemic educational policy and delivery in Europe. *Policy and Society*, *41*(1), 111–28.
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*, 75–89.
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.
Zuboff, S. (2022). Surveillance capitalism or democracy? The death match of institutional orders and the politics of knowledge in our information civilization. *Organization Theory*, *3*(3).
Zürn, M. (2018). *A theory of global governance: Authority, legitimacy, and contestation*. Oxford University Press.

Index