



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Universitat Autònoma de Barcelona

PhD programme in Sociology

Department of Sociology

Faculty of Political Science and Sociology

From trust in the profession to trust in results: A multi-scalar analysis of performance-based accountability in Norwegian education

Doctoral Dissertation by
Marjolein K. Camphuijsen

Directed by
Dr. Antoni Verger Planells
Dr. Antonina Levatino

Barcelona, December 2020



Universitat Autònoma de Barcelona

PhD programme in Sociology

Department of Sociology

Faculty of Political Science and Sociology

From trust in the profession to trust in results: A multi-scalar analysis of performance-based accountability in Norwegian education

Doctoral Dissertation by
Marjolein K. Camphuijsen

Directed by
Dr. Antoni Verger Planells
Dr. Antonina Levatino

External supervision by
Dr. Guri Skedsmo

Barcelona, December 2020

Ik draag dit proefschrift op aan mijn moeder, die al wist dat sociologie iets voor mij was voordat ik
het zelf wist.

This research was supported by the European Research Council under Grant 680172.

ACKNOWLEDGEMENTS

I feel greatly indebted to many people who have contributed in different ways to the realization of this dissertation. First of all, I would also like to express my deepest gratitude to my thesis supervisors, Antoni Verger, Antonina Levatino and Guri Skedsmo.

Toni, I cannot thank you enough for the opportunity to conduct my PhD research as part of the 'REFORMED' research project. It has been a fantastic and hugely stimulating experience to be part of such an inspiring team of critical scholars. Thank you also for your patience and interest in the questions and themes I have wished to explore, your continuous support throughout the process, and your essential comments and constructive feedback.

Anto, I lack the words for how grateful I am for your encouragement, help and extensive feedback. I have learned so much from working so closely together and from our many inspiring conversations. Thank you also for always lending me your ear and for your emotional support during difficult times. It has been crucial. I hope we get many chances to open our "freezer of ideas" in the future.

Guri, thank you so much for your guidance and advice throughout the dissertation process. Your support, expertise and insights have been invaluable. Talking to you always helps me to clarify my ideas, and has been of immense value in strengthening the dissertation.

I am also deeply grateful for the support provided by Jorunn Møller throughout the research process. Jorunn, thank you so much for your help during the first round of data collection, for always patiently reading my work, for giving me that push when I needed it, and for telling me it is enough when it is. Thank you, and Guri, also for your patience in listening to me practice my Norwegian. Your encouragement has given me the confidence I needed, and has been essential for the data collection of this dissertation.

I cannot express enough gratitude to all research participants, who put time and effort in answering my questions, sharing their thoughts, and telling me about their experiences. Without them, this dissertation in its current form would not have been possible.

This research reported on in this dissertation has been supported by the European Research Council under Grant 680172, awarded to the 'REFORMED' project.

I have furthermore benefited greatly from the support and insights from a number of scholars, colleagues and friends, who have helped me shape this study. The numerous interactions and collaborations with team members of the 'REFORMED' project have been crucial. Andreu, Anto, Clara, Edgar, Gerard, Giulia, Isabel, Laura, Lluís, Marcel, Mauro, Natalie, and Toni, thank you all so much for creating and contributing to such a stimulating and warm research environment. Your enthusiasm for the research topics has been infectious. This dissertation has benefited greatly from your critical questions, valuable insights, and constructive feedback throughout all phases of the research. Thank you also for your kindness and the many good times shared.

A special thanks to Natalie, with whom I shared this research journey most intensively. Conducting a PhD is often described as a solitary pursuit, but sharing this experience with you has meant that it rarely felt like that. Thank you also for your positive energy and trust in good endings.

Thank you Toni, Anto, Guri, Lluís and Clara for all your extra help during the final stage of this dissertation. The final months have been the hardest of all, and I doubt I could have done it without you.

I would furthermore like to thank all members of the research group 'Globalisation, Social Policies and Education' at the Autonomous University of Barcelona. I have greatly appreciated the intellectually stimulating research environment, and thank you all for your generosity in making me feel at home in Barcelona.

I have also felt fortunate to have met members of the 'New AGE' research project at several occasions. Their work on 'soft' and 'reflexive' accountability has been a great source of inspiration and help throughout my PhD.

I am grateful for the warm welcome and continuous support provided by the Department of Teacher Education and School Research and the Department of Education at the University of Oslo. It has been a great pleasure to take part in the research group 'Curriculum Studies, Leadership and Educational Governance'. Ann Elisabeth, Berit, Eivind, Eli, Ida, Jeff, Jorunn, Guri, Kirsten, Maike, thank you all for your kindness and your local insights.

A special thanks to Eivind. I greatly enjoyed our many talks during my visits to Oslo, and will surely miss these.

I would also like to thank Melanie Ehren for our inspiring conversations on accountability and school governance reforms. Your lecture on the sociology of quantification greatly helped me to make sense of my data. Its timing could not have been more perfect. Thank you also for the opportunity to join the research group 'Educational Governance, Identity and Diversity' for a research stay. I feel very fortunate to continue my research journey at the Free University of Amsterdam, and very much look forward to our future collaboration.

I feel furthermore indebted to Ann Elisabeth Gunnulfsen, Sapideh Sadeghi and Anja Johansen for their support during the administration of the 'REFORMED' questionnaire in Norway. Without your local insights, your persistence and enthusiasm, we would never have received so many filled-out questionnaires.

Lastly, I would like to acknowledge my friends and family. I could not have finished this dissertation without their unconditional support and encouragement. Diego, thank you for being the best partner to share this journey with.

ABSTRACT

In an attempt to raise the performance, efficiency and equity of education systems, performance-based accountability policies have been adopted or strengthened by an increasing number of countries, characterized by highly diverse political institutional regimes, education systems, and levels of economic development. While research interest in accountability reforms has sparked in recent decades, the existing literature remains characterized by an insufficient understanding of why policymakers, in such diverse contexts, turn towards performance-based accountability to reform their education systems, in particular considering the absence of strong evidence regarding the benefits produced by this reform approach. In a similar vein, a limited understanding prevails of the mechanisms and conditions under which performance-based accountability generates particular policy outcomes.

With the aim of contributing to filling some of the gaps in the existing literature, this doctoral dissertation adopts a multi-scalar case-study, which scrutinizes the different (but interconnected) scales in the deployment of performance-based accountability policies in Norway, a Nordic country formerly considered 'immune' to this globalizing policy phenomenon. More specifically, the dissertation examines how performance-based accountability, as a global policy for education reform, is adopted, shaped and enacted in the Norwegian education system. In order to examine the trajectory of performance-based accountability in Norway, the analysis relies on two principal data sources; 37 in-depth interviews with top-level politicians, policymakers and education stakeholders, as well as 4 White Papers. Moreover, to shed light on the role played by intermediate bodies in mediating the ways in which accountability plays out in practice, a systematic analysis of 3,046 newspaper articles published by 155 Norwegian regional and local newspapers between 2004-2018 is performed. Finally, to gain a deeper understanding of how accountability demands are interpreted, experienced and put into practice at the school level, the analysis relies on in-depth interviews with 23 school principals working at public, primary schools in nine urban municipalities.

The findings highlight how performance-based accountability has been adopted and institutionalized in Norway as a way to secure equity and quality standards in a highly decentralized education system. While global drivers and policy ideas have clearly influenced policy design processes, the trajectory of performance-based accountability in Norway is simultaneously shaped by the country's political institutions, welfare regime, and education system, as well as by policy enactment processes. The media

play a key role in mediating the impact of the publication of test results as a key pressure mechanism to elicit the desired behavioral change. The analysis identifies four dominant frames present in regional and local media coverage on standardized testing, as well as important differences in media reporting over time and across localities. Finally, the findings identify distinct response patterns in how Norwegian principals perceive, interpret and translate accountability demands, which range from alignment to accountability expectations to symbolic responses.

By highlighting how performance-based accountability can drive behavioral change by reframing norms of good educational practice, and by changing how educators make sense of core aspects of their work, this dissertation challenges a central premise of educational accountability research; i.e. that external incentives and sanctions form the prime explanatory factors of behavioral change instigated by accountability reforms. Moreover, the dissertation contributes to the educational accountability literature by providing novel empirical findings on a lesser documented accountability regime in a Nordic educational setting, as well as by demonstrating the value of a multi-scalar approach to understand the deployment of accountability policies. The dissertation advocates that lowering the stakes forms an insufficient measure to prevent perverse effects from occurring, and calls for the need to re-think current dominant accountability approaches in education.

LIST OF ABBREVIATIONS

CPE	Cultural Political Economy
MER	Ministry of Education and Research
NPM	New Public Management
NQAS	National Quality Assessment System
OECD	Organisation for Economic Co-operation and Development
SES	Socio-Economic Status
PBA	Performance-Based Accountability
PISA	Programme for International Students Assessment
PIRLS	Progress in International Reading Literacy Study
TBA	Test-Based Accountability
TIMMS	Trends in International Mathematics and Science Study
WP	White Paper

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	6
ABSTRACT.....	9
LIST OF ABBREVIATIONS	11
CHAPTER 1: INTRODUCTION.....	15
Previous research on PBA policies in education: main contributions and gaps	17
Understanding the global popularity of PBA	17
Understanding the policy outcomes of PBA.....	19
Gaps in our understanding of the deployment of PBA policies	20
Purpose of the dissertation	22
Contextual background.....	23
Research questions	26
Methodology	27
Dissertation structure	30
References	30
Compendium of publications	38
CHAPTER 2 TEST-BASED ACCOUNTABILITY IN THE NORWEGIAN CONTEXT: EXPLORING DRIVERS, EXPECTATIONS AND STRATEGIES	40
Introduction	40
Theoretical approach and analytical concepts	42
The Norwegian educational context	43
Data and methodology	44
Findings	46
Variation: From ‘best school in the world’ to ‘school loser’	47
Selection: ‘the most research-based reform’	48
Retention: ‘inclusion as a basic ethical warrant’.....	52
Discussion and conclusion	55
Notes	58
References	59
CHAPTER 3: SCHOOLS IN THE MEDIA: FRAMING NATIONAL STANDARDIZED TESTING IN THE NORWEGIAN PRESS, 2004-2008	65

Introduction	65
National testing in Norway	66
Examining media coverage on standardized testing through frame analysis	67
Data and methodology	68
Framing national standardized testing in the Norwegian press.....	69
Four most recurrent frames: evolution over time and geographical differences	70
The frame of ‘performance’	75
The frame of ‘transparency and empowerment’	76
The frame of ‘misinterpretation and misuse’	78
The frame of ‘criticism’	79
Discussion	80
References	81
CHAPTER 4: COPING WITH PERFORMANCE EXPECTATIONS: TOWARDS A DEEPER UNDERSTANDING OF VARIATION IN SCHOOL PRINCIPALS’ RESPONSES TO ACCOUNTABILITY DEMANDS	89
Introduction	89
Literature review: school actors’ responses to PBA.....	92
Reactivity as an analytical device to interpret and explain responses to PBA	93
Measurement, transparency and accountability in the Norwegian context.....	96
Data and methodology	97
Findings	100
Alignment: Achievement of basic skills as core school mission	103
Balancing multiple purposes: Performance data to inform decision-making	107
Symbolic responses: Value-driven education and a focus on the non-measurable.....	110
Discussion and conclusion	114
References	117
CHAPTER 5: SUMMARY OF MAIN RESEARCH FINDINGS.....	122
References	130
CHAPTER 6: CONCLUSIONS.....	132
Methodological insights	132
The value of a multi-scalar approach	133
Accountability in a Nordic education setting: Learning from the Norwegian experience.....	135

Understanding the reactive interactions between actors and measures 137
Limitations and future lines of inquiry 138
Policy relevance 140
References 142

CHAPTER 1: INTRODUCTION

In recent decades, concerns about the quality, efficiency and equity of education systems have generated pressure on national policymakers to reform education. Policy principles such as decentralization, learning standards, and accountability have taken a central place in education reforms worldwide. Increasingly, these policy principles are integrated into a compact reform approach by means of large-scale standardized evaluation instruments and students' performance metrics (Ball, Junemann, & Santori, 2017; Verger, Parcerisa, & Fontdevila, 2019). On the one hand, this reform approach tends to imply a devolution of organizational and pedagogical responsibilities to lower government levels and schools, while on the other hand, local authorities and school actors are increasingly subject to performance monitoring and held accountable for the extent to which they obtain centrally-defined objectives, usually measured by standardized tests. This form of accountability, conceptualized as performance-based accountability, has spread globally, and is adopted by an increasing number of countries characterized by highly diverse political institutional regimes, education systems, and levels of economic development (Ball, Junemann, & Santori, 2017; Kamens & McNeely, 2010; Lingard, Martino, Rezai-Rashti, & Sellar, 2015; Verger, Parcerisa, & Fontdevila, 2019).

The global popularity of performance-based accountability (hereafter PBA) is noteworthy, considering that research on the operation and impact of PBA remains characterized by weak, inconclusive and contradictory effects on student learning outcomes, instructional improvement and education inequalities (e.g. see Falabella, 2014; Jarke & Breiter, 2020; Maroy, 2015; Maroy & Pons, 2019; Sahlberg, 2016; Verger & Parcerisa, 2017). It has moreover been documented how PBA policies, especially when high stakes are attached to them, can generate a number of perverse or undesired effects, such as curriculum narrowing (Berliner, 2011) and teaching to the test (Au, 2007), as well as opportunistic behavior including 'cream-skimming' (Jennings, 2010), educational triage (Booher-Jennings, 2005; Reback, 2008) and cheating (Jacob & Levitt, 2003). Finally, studies have documented a range of negative effects on the teaching profession, including reduced job satisfaction (Pearson & Moomaw, 2005), reduced motivation (Day, 2002), and increased stress and burnout (Berryhill, Linney, & Fromewick, 2009; Dworkin, Saha, & Hill, 2003; Perryman, Ball, Maguire, & Braun, 2011).

Against this backdrop, a growing body of research has examined *why* PBA policies are spreading globally, despite there being weak evidence on the benefits generated by this reform approach.

Simultaneously, a second body of research has studied the varied policy responses adopted in reaction to PBA demands, which range from accommodation of accountability expectations to symbolic responses, negotiation and resistance (e.g. Barbana, Dumay, & Dupriez, 2019; Candido, 2020; Diamond & Spillane, 2004; Maroy & Pons 2019; Verger, Ferrer-Esteban, & Parcerisa, 2020). Nonetheless, despite important contributions, the existing literature on PBA remains characterized by a number of gaps, which limit the understanding of why policymakers, in such diverse contexts, turn towards PBA to reform their education systems, as well as the mechanisms and conditions under which PBA generates particular policy responses and outcomes.

The starting point of this dissertation is that, in order to gain a deeper and more comprehensive understanding of why PBA policies are spreading and how they are generating particular policy outcomes, it is necessary to extend empirical coverage of under-researched accountability approaches, conduct multi-scalar analyses of the deployment of accountability policies, and examine the role played by third-party account-holders, such as for example the media, in mediating and conditioning policy enactment processes. To do so, this dissertation adopts a multi-scalar case-study, which scrutinizes the different (but interconnected) scales in the deployment of accountability policies in Norway, a Nordic country formerly considered ‘immune’ to this globalizing policy phenomenon (Verger, Parcerisa, & Fontdevila, 2019). More specifically, the multi-scalar approach translates into three sub-studies. By relying predominantly on qualitative research methods, the sub-studies examine the trajectory of PBA in Norwegian education, its mediation by intermediate bodies, and enactment at the school level. In so doing, this dissertation aims to provide an answer to the question how performance-based accountability, as a global policy for education reform, is adopted, shaped and enacted in the Norwegian education system.

In this introductory chapter, I will first situate the study in the broader educational accountability literature, by briefly reviewing the main contributions of research on PBA reforms, as well as by outlining the main gaps by which this body of literature is characterized. I will then describe how this dissertation contributes to filling some of these gaps. Subsequently, the rationale for conducting this research in Norway is explained, which is followed by a presentation of the research questions and methodology. This introductory chapter ends with an overview of the structure of this dissertation.

Previous research on PBA policies in education: main contributions and gaps

In general terms, accountability refers to ‘the processes, mechanisms, and/or instruments that make institutions (but also individuals and groups within institutions) meet their obligations and become more responsive to their particular audiences’ (Verger & Parcerisa, 2017, p. 450). More specifically, accountability can be defined as ‘a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences’ (Bovens, 2007, p. 450). While the relationships and purposes of accountability are inevitably of an evaluative nature, the specific criteria and judgement depend on the mode adopted (Ranson, 2003). Over the past decades, scholars have referred to an important shift in the governance of education, which goes from professional to performance-based accountability systems (Carnoy & Loeb, 2002), also conceptualized as test-based accountability (Hamilton et al., 2002) or performative accountability (Lingard, Sellar, & Lewis 2017).

Understanding the global popularity of PBA

The global popularity of PBA has sparked significant research interest. One body of research has identified various drivers of a different nature (e.g. technological, political, economic) behind the advance of PBA as a globalizing policy phenomenon (Verger & Parcerisa, 2018). For example, scholars have pointed towards the emergence of knowledge economies and the economy’s need for particular skills (in which literacy and numeracy are instrumental), as well as economic reform pressured faced by policymakers to raise the effectiveness and efficiency of education systems (Brown, Lauder, & Ashton, 2008). Others have brought to the attention the rise of a global education industry with strong incentives to promote standardized tests (Verger, Lubienski, & Steiner-Khamsi, 2016), as well as methodological and technological developments in psychometrics, which enable the increasing use of PBA instruments (Gorur, 2013; Lingard, Martino, Rezai-Rashti, & Sellar, 2016). It has also been argued that globalization has implied that the content and process of national policymaking is increasingly influenced by globalizing ideologies and policy paradigms (Ball, 1998; Dale, 1999; Rizvi & Lingard, 2010; Sahlberg, 2016; Whitty, Power, & Halpin, 1998; van Zanten, 2002). In this light, scholars have pointed towards the increasing global popularity of education reform packages that rely on the use of accountability instruments (Verger & Parcerisa, 2018), and the key role played by international organizations such as the Organisation for Economic Co-operation and Development

(OECD) in disseminating, promoting and legitimating such reform packages (Parcerisa, Fontdevila, & Verger, 2019). Simultaneously, international organizations such as the OECD play a key role in fostering an ‘educational race’ between countries through international comparisons and rankings, generating fear among policymakers to fall behind (Sellar, Thompson, & Rutkowski, 2017).

Furthermore, studies have highlighted how the expansion of PBA is related to the crisis of the welfare state and the global spread of New Public Management (NPM) as a policy paradigm (Verger & Parcerisa, 2018). Under NPM, the public sector is restructured according to managerial principles, instruments and organizational models derived from the private sector, in an attempt to modernize and raise the performance and efficiency of public service organizations (Hood, 1991; 1995). While the specific ways in which NPM is consolidated in particular settings differs according to institutional regimes and administrative cultures (Gunter, Hall, Serpieri, & Grimaldi, 2016), transparency and accountability form typical instruments of NPM reforms in many contexts, in addition to belief in managers (e.g. Hood 1991; 1995; Pollitt & Bouckaert, 2004). The consolidation of NPM is simultaneously indicative of more general changes in public governance, which have been conceptualized as a shift from ‘government control’ to ‘governance by results’ (Osborne & Gaebler, 1992; Newman, 2001; 2005) or ‘governance by numbers’ (Rose, 1991; Ozga, 2009). In other words, the state has increasingly moved away from governing by means of input-regulations, prescriptions and interventions, and towards governing by means of central objectives, performance monitoring and accountability (Ball & Junemann, 2012) and key actors in the system, *managers*, who collect and use data to monitor, evaluate and support decisions and solve problems (Gunter, et al., 2016).

While the studies reviewed above contribute to a deeper understanding of the global influences and drivers behind the advance of PBA as a global policy trend, they do not always capture the context-specific rationales and pressures behind the adoption of PBA, nor the ‘concrete and progressive implementation in the systems concerned’ (Maroy & Pons, 2019, p. 79), which often involves significant re-contextualization and translation of global policy ideas (Lingard, 2006). Consequently, it has been argued that, capturing how and why PBA policies are adopted in particular settings is important, as it allows for a more comprehensive understanding of the drivers behind the advance of PBA, as well as the specific ways in which PBA policies are recontextualized and translated. That is to say, global policy ideas and discourses tend to mingle with local history, culture, and politics, while being mediated by political institutions (Lingard, 2006), the socio-economic context of the country and path dependencies (Streeck & Thelen, 2005). As a result, a single model of PBA does not exist.

Rather, different approaches have been adopted in different settings, which vary significantly, including with regards to the stakes attached, the specific policy tools and action theory involved, and the level of alignment of policy tools (Gunter et al., 2016; Maroy, 2015; Maroy & Pons, 2019).

Understanding the policy outcomes of PBA

Beyond examining the global spread of PBA policies, studies have focused on the policy responses and outcomes associated with PBA reforms. A number of such studies have analyzed the impact of different accountability systems and instruments on student learning outcomes (e.g. Carnoy & Loeb, 2002; Hanushek & Raymond, 2005; Wößmann, 2007). While this strand of research highlights the existence of variation in policy outcomes generated by PBA instruments, these studies contribute in a limited way to the understanding of the processes and mechanisms by which PBA policies produce particular changes in schools and classrooms, and as such generate particular outcomes (Maroy & Pons, 2019). In response, a growing body of research has focused specifically on unravelling the processes and mechanisms that bring about particular policy outcomes.

The starting point of this body of research is that policy implementation is never a matter of direct and straightforward translation of abstract policy texts into practice. Rather, responding to policy demands is understood as a non-linear, negotiated and constrained social process, whereby a range of different policy ‘enactors’ operating at different scales give meaning to and translate policy demands in complex and creative ways, thereby shaping policy outcomes (Ball, Maguire, & Braun, 2012). Indeed, a growing number of studies has documented how the ways in which educators and local authorities perceive, interpret and respond to accountability demands is far from uniform. For example, whereas some educators adapt their practices in an attempt to meet accountability expectations, others respond by ignoring, avoiding, resisting or re-writing policy demands (e.g. Barbana, Dumay, & Dupriez, 2019; Candido, 2020; Diamond & Spillane, 2004; Maroy & Pons 2019; Verger, Ferrer-Esteban, & Parcerisa, 2020).

In an attempt to make sense of and explain variation in policy responses, studies have pointed both to policy design characteristics (Holloway, Sørensen, & Verger, 2017; Lingard, Sellar, & Lewis, 2017; Maroy, 2015), as well as to policy enactors’ pre-existing knowledge, beliefs and practices (Coburn, 2001; 2004; Spillane & Jennings, 1997) and organizational and social contexts (Braun, Ball & Maguire, 2011). An important role has moreover been attributed to local operating bodies, such as local

authorities or school boards, who interpret, translate and co-construct central policies, and thereby shape how accountability demands reach schools (e.g. Maroy & Pons, 2019). Finally, a small body of research has emphasized the key role played by ‘third-party’ account-holders (Busuioc & Lodge, 2016), such as the media, in influencing policy enactment processes. That is, by reporting on performance indicators and by holding those perceived as responsible accountable, media outlets can influence public perceptions, opinions and actions (Bjordal, 2016; Stack & Boler, 2007), as well as students’ and teachers’ experiences (Elstad, 2009; Lingard & Sellar, 2013). As a result, media outlets are increasingly recognized as an integral part of accountability systems (Baroutsis, 2016; Lingard & Rawolle, 2004).

Gaps in our understanding of the deployment of PBA policies

From the above, it can be derived that existing research on the deployment of PBA policies has made important progress in our understanding of the global drivers behind the spread of PBA, as well as in documenting and explaining variation in policy outcomes. Nonetheless, despite numerous studies, the educational accountability literature remains characterized by important gaps, which limit our understanding of why policymakers, in such diverse contexts, turn towards PBA to reform their education systems, as well as the mechanisms and conditions under which PBA systems generate particular effects.

More specifically, a first main gap in the existing literature on PBA relates to the unbalanced focus on ‘early-adopters’ of PBA reforms, such as the US, the UK, and Chile (Maroy & Pons, 2019; Mausethagen, 2013a; Mausethagen, Prøitz, & Skedsmo, 2017; Verger & Parcerisa, 2017). In these contexts, accountability systems have usually been developed by relying on external and closely aligned policy tools (Maroy, 2015), while attaching high stakes, such as financial rewards or sanctions, decisions affecting educators’ careers, or public rankings (Au, 2007). Moreover, accountability policies have often been combined with high levels of school choice and marketization, creating a scenario where ‘managers and teachers continuously have to compete, marketize and perform ‘successfully’ within a prevailing competition-led schema’ (Falabella, 2014, p. 3). As a result of extensive research in these settings, our understanding of the drivers and rationales behind the adoption of PBA, as well as the policy outcomes associated with PBA, is disproportionately based on findings from such accountability regimes. Nonetheless, despite an overlap in generic features, PBA policies have been adopted and strengthened in distinct ways in different contexts (Maroy, 2015; Maroy & Pons, 2019).

As such, in order to gain a more comprehensive understanding of why countries are turning towards PBA to reform their education systems, as well as how PBA policies are operationalized and re-contextualized in specific settings, and with what effects, there exists a need to extend empirical coverage of lesser-researched accountability approaches, as for example found in the Nordic countries.

The second gap in the existing literature relates to the scarcity of multi-scalar analyses of educational reforms. That is to say, existing studies on PBA reforms have often situated their analysis ‘at a particular scale of public action’ (Maroy & Pons, 2019, p. 9), such as a particular actor (e.g. teachers) or a particular level of governance (e.g. national, local or school). At best, attention is paid to other actors or levels by means of a literature review or a historical and institutional contextualization. Nonetheless, as argued by Maroy and Pons (2019, p. 9), ‘these two elements do not always allow us to readily envision the processes of negotiation, co-construction, and translation of public action at play at other levels, nor even to envisage their interactions and effects on the level of particular interest in the analysis’. By conducting a multi-scalar analysis, one is enabled to capture different dimensions of the policy process, while being able to examine how different scales interact and influence each other. This is crucial, since there is significant reason to assume that the deployment of accountability policies does not occur in a top-down manner, that the meaning-making processes of actors operating at different scales play a key role in shaping both policy trajectories and policy enactment (Maroy & Pons, 2019), and that processes of vernacular globalization are at play (Lingard, 2006). In other words, multi-scalar analyses are needed to enhance the understanding of how policy design, policy enactment and contextual contingencies interact in complex and multiple ways, so to gain both a deeper understanding of policy trajectories as well as of how, why and under what circumstances PBA policies produce particular policy outcomes.

The final gap in the existing literature relates to the observation that too often, studies pay insufficient attention to how contextual factors, including meso-level variables, mediate policy enactment processes (Holloway et al., 2017; Verger & Parcerisa, 2017). While a growing body of research recognizes that it is important to examine how school characteristics and institutional and socio-economic settings shape enactment processes, relatively little attention has been paid to the role of ‘third-party’ account-holders (Busuioc & Lodge, 2016), in particular the media. That is, while media outlets are increasingly recognized as an integral part of accountability systems (Baroutsis, 2016; Lingard & Rawolle, 2004), a limited understanding persists of what media coverage on performance indicators looks like, and how such coverage has evolved over time and across geographical localities.

Since recent evidence suggests that pressure from the media is not felt equally across territories (REFORMED survey data-base)¹, there might be reason to assume that regional and local press coverage of standardized testing differs significantly from place to place. As such, there exists a need for a deeper understanding of how local media outlets report on performance indicators, thereby potentially mediating how PBA demands are experienced by school actors.

Purpose of the dissertation

In an attempt to fill some of the gaps outlined above, this dissertation conducts a multi-scalar analysis of the trajectory of PBA in Norwegian education, its mediation by intermediate bodies, and enactment at the school level.

More specifically, this dissertation examines the rationales, processes, and actors involved in the adoption and institutionalization of PBA in Norway over a period of 15 years (2003-2018) and during three political cabinets. It furthermore unpacks the ways in which PBA, as a global policy of education reform, has been operationalized by means of different policy instruments and institutional arrangements, and consolidated and extended within the social-democratic institutional regime. In so doing, the drivers behind the spread and adoption of PBA in Norway in the political, economic and cultural domains are examined, as well as the ways in which PBA has been developed and retained by a selection of policy instruments and tools. Specific attention is paid to how societal and cultural factors, political institutions, and actors operating at different scales mediate the adoption and recontextualization of PBA in Norway.

In addition, this dissertation examines the role of regional and local media outlets as ‘third-party’ account-holders (Busuioc & Lodge, 2016), by performing a systematic analysis of Norwegian regional and local press coverage of standardized testing, since the adoption of the tests in 2004, until 2018. In so doing, it aims to identify the most recurrent discursive frames used by the media when informing readers about national standardized tests, as well as their main features, and evolution over time. Moreover, by focusing specifically on the regional and local press, rather than the national press, it is explored whether geographical differences exist in how media outlets report on and frame

¹ Forthcoming in www.reformedproject.eu

standardized testing, which could potentially mediate how PBA demands are experienced and enacted at the school level.

Finally, this dissertation opens up the ‘black box’ of policy enactment by examining how accountability demands are interpreted, experienced and put into practice at the school level in different local contexts. In so doing, specific attention is paid to how the process of policy enactment is mediated by actors’ pre-existing knowledge, beliefs and practices, school specific factors, as well as meso-level variables related to the local regulatory context and level of involvement of ‘third-party’ account-holders.

By examining the deployment of accountability policies at different scales of public action, this dissertation aims to illuminate how actors, processes and mechanisms at different scales shape the trajectory and enactment of PBA policies. In so doing, a more complete picture is provided of how policy design, policy enactment and contextual contingencies interact in complex ways. This way, a deeper understanding is gained of why, how and under what circumstances particular policy outcomes are generated. Moreover, by examining the adoption, operationalization and enactment of PBA in Norwegian education, this dissertation aims to provide novel empirical evidence on a lesser documented accountability regime. This way, a more comprehensive understanding is gained of why countries around the world, characterized by diverse welfare-regimes, administrative traditions, and levels of economic development, are turning towards PBA to reform their education systems, as well as how PBA policies are operationalized and re-contextualized in specific settings, and with what effects. Nonetheless, beyond the fact that Norway forms a lesser documented case in the educational accountability literature, there are three specific reasons for why Norway was chosen for this object of study, which are now explained in more detail.

Contextual background

The first reason why Norway forms a particularly relevant case for this object of study relates to the transformation in school governance that has occurred in Norway over the past decades (Møller & Skedsmo, 2013). Historically, the education profession enjoyed significant trust and substantial autonomy to exercise discretion in their work and manage their own practices (Mausethagen, 2013b). The education system was regulated by means of the Education Act and curriculum guidelines, formulated in the form of broad national aims and content, thereby allowing significant interpretive

leeway for educators to develop locally adapted teaching programs. Investing in teacher education formed a key strategy to guarantee education quality (Werler & Sivesind, 2007), and teachers and school principals were not subject to external performance monitoring of the outcomes of teaching (Hopmann, 2003; Sivesind & Bachmann 2008; Skedsmo, 2011; Werler & Sivesind, 2007). Nonetheless, over the past decades, a shift from input- to output-management has occurred (Skedsmo, 2009; Aasen, Prøitz, & Sandberg, 2014).

Since the late 1980s, educational reforms in Norway have been influenced by managerial ideas such as NPM. In particular targeted at the local school administration at the municipal level, NPM reforms promoted decentralization, deregulation, horizontal specialization and management by objectives (Christensen & Læg Reid, 2011). Moreover, in the aftermath of the publication of the first Program for International Student Assessment (PISA), students' academic achievements started to dominate the public debate on education (Møller & Skedsmo, 2013). Increasingly, concerns about below-expected learning outcomes of Norwegian students in basic skills such as reading contributed to the call for external assessment of student performance as well as external control of educators' competence and results (Møller & Skedsmo, 2013). In 2004, a national quality assessment system was adopted, which includes national standardized tests. Two years later, as part of the 'Knowledge Promotion' reform, an outcome-based curriculum was introduced. While central authorities are in charge of defining objectives and learning outcomes, local and school actors are responsible for deciding upon content and working methods in order to reach those objectives. National tests have become a central steering device to assess to what extent schools and local authorities meet central objectives (Aasen et al., 2014). The increasing use of quality assessment measures such as standardized tests and value-added models to hold teachers, school leaders and local authorities accountable for student learning outcomes has led scholars to refer to a shift from trust in the profession to trust in results (Uljen, Møller, Årlestig, & Frederiksen, 2013).

The shift from input- to output-governing, and the strong emphasis of recent policies on performance monitoring, accountability, and expectations around data use to improve educational practices (Skedsmo & Møller, 2016), is noteworthy, considering that standardized testing and teacher monitoring and evaluation were once considered controversial and out of step with Norwegian values and traditions. This makes Norway a particularly appropriate case to explore how and why PBA started to dominate educational debates, as well as to examine the rationales, processes and actors involved in promoting PBA as a suitable and desirable policy solution.

The second reason why Norway was considered relevant for the object of research relates to the ways in which PBA has been operationalized and institutionalized in the social-democratic institutional regime. As documented by previous research (e.g. Langfeldt, Elstad, & Hopmann, 2008), as well as extended in this dissertation, the Norwegian way of PBA differs in important regards from ‘hard’ or ‘strict’ accountability approaches (cf. Maroy, 2015), developed by ‘early adopters’ such as the US, the UK, and Chile. More specifically, the Norwegian approach to PBA is characterized by (1) a relative absence of accountability consequences of a material nature (such as financial rewards or sanctions; or decisions affecting schools or educators’ careers) (Langfeldt, Elstad, & Hopmann, 2008); (2) low levels of marketization and restricted school choice; and (3) a combination of external control devices as well as institutional arrangements that encourage self-evaluation processes and target internal feelings of responsibility and reflection (Camphuijsen, Møller, & Skedsmo, 2020). As such, at least at the central level, Norway seems to form an example of a ‘reflexive’ PBA system, informed by the conception of a socially embedded and reflective actor, while relying on a combination of external and internal devices to reorient actors’ behavior (cf. Maroy, 2015). While ‘hard’ accountability approaches, found in the Anglo-Saxon countries as well as in Chile, have been object of extensive research, relatively little is known about how other PBA approaches, including ‘reflexive’ and ‘soft’ approaches, play out in practice (Mausethagen, 2013a; Mausethagen, Proitz, & Skedsmo, 2017; Maroy & Pons, 2019).

The third and final reason why Norway is a particularly interesting case from an analytical point of view relates to the existence of significant local variation in how PBA is mediated by intermediate actors. The local education authority at the municipal level constitutes the formal account-holder determining how schools in the municipality are monitored and controlled. Local authorities are obliged to establish a quality assessment system, containing routines and procedures for following-up schools based on quality assessment measures, including national tests. As shown by previous research (e.g. Seland, Vibe, & Hovdhaugen, 2013; Skedsmo, 2018; Proitz, Mausethagen, & Skedsmo, 2020), significant local variation is found in local accountability regimes, following municipal discretion in terms of accountability tools and consequences. On top of this, local differences are found in the role played by external audiences, or ‘third-party’ account-holders, such as parents (Busuioc & Lodge, 2016). That is, while school choice is generally restricted in Norway at the level of compulsory education, local exceptions are found. A handful of municipalities, largely those more urbanized, have

adopted more flexible regulations, allowing families some degree of school choice² (Bjordal, 2016; Haugen, 2019; NSD, 2016). Such local variability makes Norway an excellent case to advance the understanding of how different accountability configurations and local policy contexts mediate policy enactment processes and policy outcomes.

Research questions

This dissertation seeks to examine the trajectory, mediation, and enactment of PBA policies in Norwegian education. The overarching research question of this dissertation is the following:

How is performance-based accountability, as a global policy for education reform, adopted, shaped and enacted in the Norwegian education system?

This overarching research question is examined by means of three sub-studies, each of which focuses on different scales in the deployment of accountability policies, and is guided by a specific research question:

- 1) *Why did Norway adopt PBA in education at the turn of the millennium, and how did central authorities operationalize and develop PBA through different policy instruments and tools over a period of 15 years (2003-2018)?*
- 2) *What are the most recurrent discursive frames used by the Norwegian regional and local press when informing readers about national standardized tests, and to what extent does the pervasiveness of particular frames differ over time and across geographical localities?*
- 3) *How do Norwegian school principals reflect on and respond to PBA demands, and how can variation in principals' response patterns be explained?*

² Data from the Norwegian Centre for Research Data (NSD) reveal that 56 Norwegian municipalities (i.e. 19.5%) allow for 'freer user choice in the area of compulsory education' (NSD, 2016, p.256). Still, while such municipal regulations are likely to award families some degree of school choice, priority remains given to students residing in the school's catchment area. The data are based on "Survey on Municipal Organization 2016, Municipality File". The survey was financed by the Ministry of Local Government and Modernisation (KMD). The data are provided by Norwegian Institute for Urban and Regional Research (NIBR), and prepared and made available by NSD – Norwegian Centre for Research Data. Neither NIBR, KMD nor NSD are responsible for the analyses/interpretation of the data presented here.

Whereas the first question focuses on policy adoption and evolution at the national and supra-national scale, the second question zooms in on the regional and local scales. Finally, the third question explores policy enactment at the school level. Nonetheless, while each sub-study focuses on different scales in the deployment of accountability policies, it is important to emphasize the assumed interconnectedness of the different scales of analysis. In other words, while the different sub-studies move from macro to meso to micro, it is key to remember that there is reason to suspect that the deployment of accountability policies does not occur in a top-down manner, but rather, that policy design, policy enactment and contextual contingencies interact in complex and multiple ways. As such, a main focus of this dissertation is to explore the value of a multi-scalar approach in shedding light on the deployment of accountability policies.

Methodology

To examine the research questions outlined above, this study employs a case-study. As the primary objectives of this study are to gain to an in-depth understanding of why PBA policies resonate with central authorities in Norway, as well as how these policies are operationalized, shaped, and enacted at different scales, a case study was considered the most suitable research design. Namely, a case-study allows for direct observation of a phenomenon within its ‘real-life’ context, thereby enabling the researcher to examine questions relating both to *what* happened, as well as to *how* or *why* something happened (Yin, 2006). To examine the research questions, the methodology of this dissertation relies predominantly on qualitative research methods, so to be able to gain an in-depth understanding of how different actors perceive, interpret and (co-)construct PBA policies, while being able to capture how contextual contingencies mediate policy design and enactment processes. Moreover, qualitative research methods are especially suitable to identify the specific mechanisms, processes and circumstances under which PBA policies generate specific policy responses and outcomes. Simultaneously, despite the predominant use of qualitative research methods, quantification of codes was performed during the second sub-study, so to be able to disclose trends over time and across geographical localities in media coverage of standardized testing.

An overview of each sub-study is provided in Table 1.

Table 1. Overview of the three sub-studies

Sub-study	Research question	Data sources	Data analysis
1	Chapter 2: Why did Norway adopt PBA in education at the turn of the millennium, and how did central authorities operationalize and develop PBA through different policy instruments and tools over a period of 15 years (2003-2018)?	4 White Papers and 37 in-depth interviews with top-level politicians, policymakers, and education stakeholders	Qualitative content analysis using Atlas.ti
2	Chapter 3: What are the most recurrent discursive frames used by the Norwegian regional and local press when informing readers about national standardized tests, and to what extent does the pervasiveness of particular frames differ over time and across geographical localities?	3,046 newspaper articles published by 155 Norwegian regional and local newspapers between 2004-2018	Quantification of codes; Qualitative content analysis
3	Chapter 4: How do Norwegian school principals reflect on and respond to PBA demands, and how can variation in principals' response patterns be explained?	23 in-depth interviews with school principals working at public, primary schools located in nine urban municipalities	Qualitative content analysis using Atlas.ti

As can be derived from Table 1, to examine the research question of the first sub-study, the analysis relied on two principal data sources: White Papers and in-depth interviews with top-level politicians, policymakers, and education stakeholders. The White Papers were selected after an initial screening of all White Papers published by the Ministry of Education and Research over the past 20 years. The selected White Papers represent key moments in the trajectory of PBA in Norway. The selection of interview participants was purposive: all participants have played a key role in education policy design, formation and/or enactment over the past three decades, are experts on recent changes in education policy, or are influential in the educational debate. Nonetheless, to avoid the risk of selection bias, the sample was expanded and refined during the fieldwork process by asking interviewees to identify other

potentially relevant participants. The interviews followed a semi-structured interview script (see Fontdevila, 2019). The policy documents and the interview transcripts were examined by conducting qualitative content analysis. During the analysis of the data, ATLAS.ti software was used.

The second sub-study explores the role played by intermediate bodies in mediating the ways in which PBA plays out in practice. More specifically, this sub-study focuses on how Norwegian regional and local newspapers report on and frame standardized testing. In so doing, this sub-study relied on a systematic analysis of newspaper articles on standardized testing, published by Norwegian regional and local newspapers since the adoption of the tests in 2004, up until 2018. All newspaper articles were coded using a codebook, which was the product of a dynamic, nonlinear approach. Subsequently, the sub-study relied on qualitative content analysis (McNamara, 2005). Moreover, by quantifying all codes, a unique data-base was constructed on regional and local press coverage on standardized testing, which allowed for disclosing trends across time and place.

The third sub-study examines policy enactment at the school level, and aims at gaining an understanding of how school principals reflect on and respond to PBA demands, and how variation in response patterns can be explained. This sub-study focused specifically on school principals, as principals play a key role in the enactment of accountability policies, mediating teacher experiences and responses (Diamond & Spillane, 2004). The analysis built on qualitative data generated through in-depth interviews with 23 school principals working at public, primary schools in nine urban municipalities. The sampling strategy of this sub-study consisted of two steps. Recognizing that local contextual factors mediate enactment processes at the school level (Braun et al. 2011), the sampling of municipalities aimed at ensuring variability in a) the level of ‘strictness’ of the local accountability regime (i.e. whether municipalities had a strong or weak performance orientation, as well as the type and level of alignment of the local accountability instruments); and b) the level of involvement of ‘third-party’ account-holders (i.e. parents and the media) (Busuioc & Lodge, 2016). With regard to the first selection criterion, the sub-study relied on previous research, local policy documents, and survey data. The latter data were collected in the context of the ‘REFORMED’ research project³. With regard to the second selection criteria, secondary data on school choice regulations were used (NSD, 2016), as well as the database on regional and local press coverage constructed as part of the second sub-study. Following this classification of municipalities, primary schools were selected in different local

³ www.reformedproject.eu

contexts. For the selection of schools, secondary data provided by the Norwegian Directorate for Education and Training, as well as survey data derived from the administration of the electronic questionnaire in the context of the 'REFORMED' research project were used. More specifically, the sampling of schools relied on the following, theoretically-relevant selection criteria: a) school performance; b) reported performative pressure levels; c) reported reputation and d) reported pressure to maintain enrolment perceived by principals (see Ferrer-Esteban, 2020). The interviews followed a semi-structured interview script, and were analyzed by conducting qualitative content analysis. During the analysis of the data, ATLAS.ti software was used.

Dissertation structure

This dissertation is structured as follows. The first sub-study is discussed in Chapter 2 (publication A). In Chapter 3, the second sub-study is discussed (publication B). Subsequently, in Chapter 4, the third sub-study is discussed (publication C). Thereafter, in Chapter 5, a summary of the main findings of this dissertation is provided, and an answer to the overarching research question is formulated. In the concluding chapter of this dissertation, Chapter 6, the methodological, empirical and theoretical contributions of this dissertation are highlighted. This chapter also includes a discussion of limitations of the dissertation and ideas for future research. Chapter 6 ends with a reflection on the importance of the findings from a policy point of view.

References

- Aasen, P., Prøitz, T. S., & Sandberg, N. (2014). Knowledge Regimes and Contradictions in Education Reforms. *Educational Policy*, 28(5), 718–738.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Ball, S. J. (1998). Big policies/small world: An introduction to international perspectives in education policy. *Comparative Education*, 34(2), 119–130.
- Ball, S. J., & Junemann, C. (2012). *Networks, New Governance and Education*. Bristol, UK: The Policy Press.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How Schools Do Policy: Policy Enactments in Secondary Schools*. London: Routledge.

- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu-net: Globalisation and Education Policy Mobility*. London: Routledge.
- Barbana, S., Dumay, X., & Dupriez, V. (2019). Local implementation of accountability instruments in the French-speaking community of Belgium. *European Educational Research Journal*, 19(2), 94–108.
- Baroutsis, A. (2016). Media accounts of school performance: Reinforcing dominant practices of accountability. *Journal of Education Policy*, 31(5), 567-582.
- Berliner, D. (2011). Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41(3), 287-302.
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too-Stress Provoking for Their Own Good? *International Journal of Education Policy and Leadership*, 4(5), 1-14.
- Bjordal, I. (2016). Markedsretting i en urban norsk skolekontekst. – Et sosialt rettferdighetsperspektiv. PhD dissertation, Norwegian University of Science and Technology. Retrieved from: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2428050>
- Booher-Jennings, J. (2005). Below the Bubble: “Educational Triage” and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231-268.
- Bovens, M. (2007). Analysing and Assessing Accountability: A Conceptual Framework. *European Law Journal*, 13(4), 447-468.
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, 32(4), 585-596.
- Brown, P., Lauder, H., & Ashton, D. (2008). Education, Globalisation and the Future of the Knowledge Economy. *European Educational Research Journal*, 7(2), 131-156.
- Busuioac, E.M., & Lodge, M. (2016). The reputational basis of public accountability. *Governance*. 29(2), 247-263.
- Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*, <https://doi.org/10.1080/02680939.2020.1739337>
- Candido, H. H. D. (2020). Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*, 29(1–2), 1–32.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational evaluation and policy analysis*, 24(4), 305-331.

- Christensen, T., & Lægheid, P. (2011). Democracy and administrative policy: contrasting elements of New Public Management (NPM) and post-NPM. *European Political Science Review*, 3(1), 125–146.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211–244.
- Dale, R. (1999). Specifying globalisation effects on national policy: A focus on the mechanisms. *Journal of Education Policy*, 14(1), 1–17.
- Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International journal of educational research*, 37(8), 677-692.
- Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *The Teachers College Record*, 106(6), 1145–1176.
- Dworkin, G. A., Saha, J. L., & Hill, N. A. (2003). Teacher burnout and perceptions of a democratic school environment. *International Education Journal*, 4(2), 108-120.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173-189.
- Elstad, E. (2012). PISA Debates and Blame Management among the Norwegian Educational Authorities: Press Coverage and Debate Intensity in the Newspapers. *Problems of Education in the 21st Century*, 48, 10–22.
- Falabella, A. (2014). The performing school: The effects of market and accountability policies. *Education Policy Analysis Archives*, 22(70), 1–26.
- Ferrer-Esteban, G. (2020). *Sampling strategy*. REFORMED Methodological Papers No. 3. Bellaterra: Universitat Autònoma de Barcelona.
- Fontdevila, C. (2019). *Researching the adoption of School Autonomy with Accountability reforms: A methodological note on country case studies*. REFORMED Methodological Papers No. 1. Bellaterra: Universitat Autònoma de Barcelona. doi: 10.5281/zenodo.3557302
- Gorur, R. (2013). My school, my market. *Discourse: Studies in the Cultural Politics of Education*, 34(2), 214-230.
- Gunter, H., Hall, D., Serpieri, R., & Grimaldi, E. (2016). *New Public Management and the Reform of Education: European Lessons for Policy and Practice*. London: Routledge.

- Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of TBA in education*. Santa Monica, CA: Rand Corporation.
- Hanushek, E. A., & Raymond, M. E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, 2(2/3), 406-415.
- Haugen, C. R. (2019). Teachers' experiences of school choice from 'marginalised' and 'privileged' public schools in Oslo. *Journal of Education Policy*, 35(1), 68-94.
- Holloway, H., Sørensen, T. B., & Verger, A. (2017). Global perspectives on high-stakes teacher accountability policies: an introduction. *Education policy analysis archives*, 25(85), 1-18.
- Hopmann, S. T. (2003) On the evaluation of curriculum reforms. *Journal of Curriculum Studies*, 35(4), 459–478.
- Hood, C. (1991) A public management for all seasons. *Public Administration*, 69(1), 3–20.
- Hood, C. (1995) Contemporary public management: a new global paradigm? *Public Policy and Administration*, 10(2), 104–117.
- Jacob, B. A., and Levitt, S. D. (2003). *Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. Working paper No. 9413. National Bureau of Economic Research. <http://www.nber.org/papers/w9413>
- Jarke, J., & Breiter, A. (2020). *The Datafication of Education*. New York: Routledge.
- Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227-247.
- 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.
- Langfeldt, G., Elstad, E., & Hopmann S. T. (2008). *Ansvarlighet i skolen*. Oslo: Cappelen Forlag.
- Lingard, B., & Sellar, S. (2013). 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, 28(5), 634-656.
- Lingard, B., & Rawolle. S. (2004). Mediatizing Educational Policy: The Journalistic Field, Science Policy, and Cross-field Effects. *Journal of Education Policy*, 19(3), 361–380.
- Lingard, B. (2006). Globalisation, The Research Imagination and Deparochialising The Study of Education. *Globalisation, Societies and Education*, 3(2), 287–302
- Lingard, B., Martino, W., Rezai-Rashti, G., & Sellar, S. (2015). *Globalizing educational accountabilities*. New York: Routledge.

- Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. In G. Noblit (Ed.), *Oxford Research Encyclopaedia of Education* (pp. 1-28). New York: Oxford University Press.
- Maroy, C. (2015). Comparing Accountability Policy Tools and Rationales: Various Ways, Various Effects? In H. G. Kotthoff & L. Klerides (Eds.), *Governing Educational Spaces: Knowledge, Teaching, and Learning in Transition* (pp. 35-56). Rotterdam: Sense Publication.
- Maroy, C., and Pons, X. (2019). *Accountability Policies in Education. A comparative and multilevel analysis in France and Quebec*. Cham: Springer.
- Mausethagen, S. (2013a). A research review of the impact of accountability policies on teachers' workplace relations. *Educational Research Review*, 9, 16–33.
- Mausethagen, S. (2013b). Talking about the test. Boundary work in primary school teachers' interactions around national testing of student performance. *Teaching and Teacher Education*, 36, 132-142.
- Mausethagen, S., Prøitz, T. S., and Skedsmo, G. (2017): Teachers' use of knowledge sources in 'result meetings': thin data and thick data use. *Teachers and Teaching*, 24(1), 37-49.
- Macnamara, J. R. (2005). Media Content Analysis: Its Uses, Benefits and Best Practice Methodology. *Asia Pacific Public Relations Journal*, 6(1), 1–34.
- Møller, J., & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Education Administration and History*, 45(4), 336-352.
- Newman, J. (2001). *Modernizing governance. New labour, policy and society*. London, Thousand Oaks, New Delhi: SAGE.
- Newman, J. (2005). *Remaking governance. Peoples, politics and the public sphere*. The Policy Press, University of Bristol.
- NSD. (2016). Kommunefil. NSD – Norsk senter for forskningsdata AS, NSD, 2016. <https://doi.org/10.18712/NSD-NSD2406-V1>.
- Osborne, D., & Gaebler, T. (1992) *Reinventing government*. Reading, MA: Addison-Wesley.
- Ozga, J. (2009). Governing education through data in England: from regulation to self evaluation. *Journal of Education Policy*, 24(2), 149-162.
- Parcerisa, L., Fontdevila, C., & Verger, V. (2019). Understanding the PISA Influence on National Education Policies: A Focus on Policy Transfer Mechanisms. In: S. Jornitz & A. Wilmers (Eds.), *International Perspectives on School Settings, Education Policy and Digital Strategies. A Transatlantic Discourse in Education Research* (pp. 185-198). Barbara Budrig Verlag.

- Pearson, L. C., & Moomaw, W. (2005). The Relationship between Teacher Autonomy and Stress, Work Satisfaction, Empowerment, and Professionalism. *Educational Research Quarterly*, 29(1), 38–54.
- Perryman, J., Ball, S. J., Maguire, M., & Braun, A. (2011). Life in the Pressure Cooker School League Tables and English and Mathematics Teachers' Responses to Accountability in a Results-Driven Era. *British Journal of Educational Studies*, 59(2), 179-195.
- Pollitt, C., & Bouckaert, G. (2011). *Continuity and change in public policy and management*. Cheltenham: Edward Elgar Publishing.
- Prøitz, T. S., Mausestagen, S., & Skedsmo, G. (2019). District administrators' governing styles in the enactment of data-use practices. *International Journal of Leadership in Education*, 1-22.
- Ranson, S. 2003. Public accountability in the age of neo-liberal governance. *Journal of Education Policy*, 18(5), 459-480.
- Rawolle, S., & Lingard, B. (2014). Mediatization and education: a sociological account. In K. Lundy (Ed.), *Mediatization of communication* (pp. 595-614). Berlin: De Gruyter Mouton.
- Reback, R. L. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5-6): 1394-1415
- Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. Abingdon, Oxon, U.K.: Routledge
- Rose, N. (1991). Governing by numbers: Figuring out democracy. *Accounting, Organizations and Society*, 16(7), 673-692
- Sahlberg, P. (2016). The Global Education Reform Movements and its Impact on Schooling. In K. Mundy, A. Green, B. Lingard & A. Verger (Eds.), *The Handbook of Global Education Policy* (pp. 128–144). Oxford: Wiley Blackwell.
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). Evaluering av nasjonale prøver som system [Evaluation of national testing as a system]. Published online: <https://www.nifu.no/publications/1025043/> Accessed 1 February 2020.
- Sellar, S., Thompson, G., & Rutkowski, D. (2017). *The Global Education Race: Taking the Measure of PISA and International Testing*. Toronto: Brush Education
- Sivesind, K., & Bachmann, K. (2008). Hva forandres med nye standarder? Krav og utfordringer med Kunnskapsløftets læreplaner. In G. Langfeldt, E. Elstad, & S. Hopmann (Eds.), *Ansvarlighet i skolen. Politiske spørsmål og pedagogiske svar* (pp. 62–93). Fagernes: Cappelen Akademisk Forlag.
- Skedsmo, G. (2009). *School Governing in Transition? Perspectives, Purposes and Perceptions of Evaluation Policy*. PhD dissertation, University of Oslo.

- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: inconsistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23(1), 5-20.
- Skedsmo, G. (2018). Comparison and Benchmarking as Governing Processes in Norwegian Schools. In L. M. Carvalho, L. Levasseur, M. Liu, R. Normand, & D. A. Oliveira (Eds.), *Education Policies and the Restructuring of the Educational Profession* (pp. 137-158). Singapore: Springer.
- Skedsmo, G. & Møller, J. (2016). Governing by new performance expectations in Norwegian Schools. In H. Gunter, E. Grimaldi, D. Hall & R. Serpieri (Eds.) *NPM and educational reform in Europe*. London: Routledge: pp. 53-65.
- Spillane, J. P., & Jennings, N. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *The Teachers College Record*, 98(3), 449–481.
- Stack, M., & Boler, M. (2007). Media and Policy in Education. *Policy Futures in Education*, 5(1), 1-16.
- Streeck, W., & Thelen, K. (2005). *Beyond Continuity: Institutional Change in Advanced Political Economies*. Oxford: Oxford University Press
- Uljens, N., Møller, J., Ärlestig, H., & Frederiksen, L. F. (2013). The Professionalisation of Nordic School Leadership. In L. Moos (Ed.), *Transnational Influences on Values and Practices in Nordic Educational Leadership. Is there a Nordic Model?* (p. 133-158). Dordrecht. Springer.
- Verger, A., Lubienski, C., & Steiner-Khamsi, G. (2016). *World Yearbook of Education 2016 The Global Education Industry*. New York: Routledge.
- Verger, A., & Parcerisa, L. (2017). *Accountability and education in the post-2015 period: International trends, enactment dynamics and socio-educational effects*. Think piece for the UNESCO GMER 2017. Paris: UNESCO.
- Verger, A., & Parcerisa, L. (2018). Test-based accountability and the rise of regulatory governance in education: A review of global drivers. In: A. Wilkins and A. Olmedo (eds.) *Education governance and social theory: Interdisciplinary approaches to research* (p. 139-158). London: Bloomsbury.
- Verger, A. Parcerisa, L. & Fontdevila, C. (2019). The Growth and Spread of National Assessments and Test-based Accountabilities: A Political Sociology of Global Education Reforms. *Educational Review*, 71(1), 5–30.
- Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020). In and out of the 'pressure cooker': Schools' varying responses to accountability and datafication. In S. Grek, C. Maroy & A. Verger (Eds.), *World Yearbook of Education 2021*. London: Routledge.

- Werler, T., & Sivesind, K. (2007). Norway. In: H. Döbert, W. Hörner, B. von Kopp, and W. Mitter (Eds.), *The Education System of Europe* (pp. 573-589). Dordrecht: Kluwer.
- Wößmann, L. (2007). International evidence on school competition, autonomy, and accountability: A review. *Peabody journal of education*, 82(2-3), 473-497.
- Whitty, G, Power, S., & Halpin, D. (1998). *Devolution and Choice in Education: The School, the State and the Market*. Buckingham, UK: Open University Press.
- Yin, R. (2006). *Case Study Research: Design and Methods*. *Applied Social Research Methods Series*. Thousand Oaks, London: Sage Publications.
- Zanten, A. van (2002). Educational change and new cleavages between head teachers, teachers and parents: global and local perspectives on the French case. *Journal of Education Policy*, 17(3), 289-304.

Compendium of publications

Below, an overview is provided of the articles that are part of the compendium of publications.

Publication A: Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*. <https://doi.org/10.1080/02680939.2020.1739337>

Publication B: Camphuijsen, M. K., & Levatino, A. (2020, under review). Schools in the media: framing national standardized testing in the Norwegian press, 2004-2018. *Discourse: Studies in the Cultural Politics of Education*.

Publication C: Camphuijsen, M. K. (2020). Coping with performance expectations: towards a deeper understanding of variation in school principals' responses to accountability demands. *Education Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-020-09344-6>

Publication A: Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*.
<https://doi.org/10.1080/02680939.2020.1739337>

CHAPTER 2

Test-based accountability in the Norwegian context: Exploring drivers, expectations and strategies

Introduction

Rising concerns about the performance, equity and efficiency of education systems have policy-makers around the world focussed on education reform processes. Policy principles such as decentralisation, standards and accountability are central to education debates worldwide and feature increasingly in coherent reform initiatives based on standardised assessments (Ball, Junemann, and Santori 2017; Verger, Parcerisa, and Fontdevila 2019). This means that while lower government levels and schools are given greater authority for organisational and pedagogical decisions, these actors are simultaneously held accountable for achievement of centrally defined objectives measured by standardised tests. Commonly referred to as test-based accountability (TBA) (Hamilton, Stecher, and Klein 2002), this near-universal trend is seen even in countries previously considered 'immune' to this globalising phenomenon, including Norway (Verger, Parcerisa, and Fontdevila 2019).

In the early 2000s, Norwegian authorities introduced standardised testing, teacher monitoring and evaluation and an outcome-based curriculum, while also promoting further devolution of responsibilities to local education authorities and schools. These policy initiatives represented a disruptive transformation within educational institutions and school governance (Hall et al. 2015). Standardised testing and teacher monitoring and evaluation were once considered controversial and out of step with Norwegian values and traditions. The radical shift in school governance and the adoption of once disputed policy measures has received significant research attention (e.g. see Hatch 2013; Hovdenak and Stray 2014; Langfeldt, Elstad, and Hopmann 2008; Tveit 2018). Studies have highlighted how major education reforms in Norway entailed a shift from input- to output governing, and how policy changes increasingly emphasise performance monitoring, accountability and data use to improve educational practices (Skedsmo 2009).

The present paper examines why and how TBA started to dominate educational debates in Norway, and how this policy has been operationalised by different policy tools. The following research questions are addressed: a) What is the policy trajectory of TBA in Norway? b) What are the main drivers and rationales for Norway's adoption of TBA? c) What are the main policy tools to adopt and

develop TBA in Norway? To examine these research questions, this study analyses key policy documents and primary data from 37 interviews with key educational reformers, legislators and stakeholders, conducted between September 2017 and February 2018.

We aim to move beyond previous research by focussing not only on the adoption of TBA in Norway, but by tracing the institutionalisation and evolution of TBA over the past 15 years (2003-2018) and during three political cabinets. In comparison with the moment of policy adoption, policy evolution often remains under-researched in policy research, despite the importance of seemingly small adaptations for how policy plays out in practice. In a similar vein, we aim to contribute to existing evidence by complementing policy document analysis, commonly used in (Norwegian) policy research, with in-depth interviews with key actors and stakeholders. The interviews proved especially fruitful to identify the often implicit or unarticulated ‘world views’ informing policy tool selection (Maroy 2015), and to gain insight into the policy process leading to the selection of formal measures. By examining the selection of policy instruments and their 15-year evolution, the present analysis aims to provide a deeper understanding of the advance of this globalising policy trend. Considering that existing evidence concerning the enactment and effects of accountability is inconclusive and contradictory (e.g. Maroy 2015; Sahlberg 2016), it seemed particularly useful to analyse the drivers of this global reform approach and its operationalisation within specific settings.

In so doing, the analysis is guided by the cultural political economy (CPE) framework (Jessop 2010), in combination with a political sociology-driven approach to policy instruments (Lascoumes and Le Galès 2007). CPE examines the drivers of the spread and adoption of global education policies, while the political sociology-driven approach to policy instruments explores how and why particular global policy configurations are developed and retained by selected policy instruments and tools. Both theoretical perspectives help investigate the mediation of the adoption and re-contextualisation of global policy ideas by societal and cultural factors and institutions, and actors operating at different scales.

In the sections below, we first briefly discuss the different nature of globalising TBA modalities, before outlining the study’s guiding theoretical approach and analytical concepts. Thereafter, we describe the Norwegian educational context and school system to provide a context for the study’s findings. Then follows a description of the data and methodology. In the section thereafter, we present our main results, structured according to the three evolutionary mechanisms identified by our theoretical model (Jessop, 2010). We conclude this paper with arguing that while TBA formed a key policy instrument

to ensure equity and quality standards in the highly decentralised Norwegian education system, to some degree both equity and quality have been rearticulated to performance indicators based on national and international tests.

Theoretical approach and analytical concepts

In recent decades, educational agendas have been influenced by globalising ideologies and policy paradigms, accompanied by reform packages based on similar discourses and rationales (Ball 1998; Sahlberg 2016). Promoted by international organisations such as the Organisation for Economic Co-operation and Development (OECD), TBA modalities are anticipated to raise the efficiency, academic excellence and equity of education systems. Although grounded in similar principles, TBA modalities around the world differ in a) degree of regulatory tool alignment; b) consequences of accountability measures; c) the conceptions of actors; and d) the nature of mediation underlying regulation (Maroy 2015). Therefore, policy trajectories and how policies are operationalised in context are important to examine.

The present analysis is informed by Jessop's (2010) CPE framework, which forms a useful heuristic device to study processes of profound institutional transformation. CPE examines economic, political and cultural factors that interact through three evolutionary mechanisms that characterise policy adoption and change: variation, selection and retention. Variation is triggered by a perceived need to revisit existing policies or practices, prompting selection of suitable and agreed policy instruments. Retention institutionalises new policy proposals and instruments in the context of existing regulations and practices. Both symbolic and material factors influence the process of policy change (Jessop 2010). These three evolutionary mechanisms provide useful analytical tools for systematic examination of the contingencies, events and actors involved in policy change, and identification of mechanisms inducing or restraining institutional change. Using this framework, we aim to contribute to a holistic explanation of why and how TBA has been adopted and developed in Norway, by identifying drivers of change in the political, economic and cultural domains.

Using a political sociology-driven account of policy instruments, we explore the operationalisation and institutionalisation of TBA over time, examining the main policy tools to adopt and develop TBA in Norway. This approach criticizes functionalist approaches that view public policy as pragmatic, policy instruments as 'natural' and at policy-makers' disposal and policy adoption as informed by instruments'

proven effectiveness. In so doing, functionalist approaches oversimplify processes of instrumentation by neglecting economic and political factors. In contrast, the political sociology-driven approach to policy instruments holds that the process of instrumentation must be further problematised (Lascoumes and Le Galès 2007). Here, public policy instruments are seen as ‘bearers of values, fuelled by one interpretation of the social and by precise notions of the mode of regulation envisaged’. We seek the political and economic stakes of instrumentation by tracing ‘power relations associated to instruments and issues of legitimacy, politicisation, or de-politicisation dynamics associated with different policy instruments’ (Lascoumes and Le Galès 2007, 4).

Combining political sociology and CPE lenses, we investigate how the adoption, development and operationalisation of these policy instruments and tools are contingent on and mediated by institutional and contextual factors. As Verger et al. (2016) showed, the economic, political, institutional and cultural factors that provoke, condition and legitimise domestic policy formation may differ widely. For example, factors such as a country’s economic environment (Lenschow, Liefferink, and Veenman 2005) or an economic crisis or recession (Ball 1990) can create pressure for reform and shape the perception of policy instruments as legitimate and economically feasible. Political factors, motivations and interests can also influence the policy problems that education reforms will address, as will the policy instruments developed for that end. Cultural factors including ‘the semiotic and meaning-making dimension of education policy processes’, and social values or public opinions can similarly shape the adoption and retention of particular policy ideas and instruments (Verger et al. 2016). Finally, our approach acknowledges that institutional legacies often mediate rather than provoke policy change. Public administration traditions, political institutions and regulatory frameworks shape policy-makers’ views about new policies and define the institutional boundaries of policy formation.

The Norwegian educational context

The Norwegian social-democratic welfare state (Esping-Andersen, 1990) has high levels of public social expenditure and direct provision of public services by state and local government. Governance responsibilities are divided between municipalities, counties and the state. Since the late 1980s, neoliberal thinking and new managerial ideas such as New Public Management (NPM) have influenced

public sector reforms in Norway, promoting decentralisation, deregulation, horizontal specialisation and management by objectives (Christensen and Læg Reid 2011).

Education is predominantly publicly provided and financed and aims to promote values such as equity, solidarity, social justice and democracy. The comprehensive school model seeks equal opportunity for people of all genders, socio-economic or ethnic backgrounds, or geographical locations (Blossing, Imsen, and Moos 2014). While 82% of Norway's population lives in urban areas (Statistics Norway 2019a), many municipalities and schools are small. School choice is limited, especially for compulsory education, although local exceptions are found (Haugen 2019). The establishment of private schools is strictly regulated, with only 4% of the school-aged population enrolled in private primary or lower-secondary schools (Statistics Norway 2019b). Norway's national curriculum guidelines were traditionally formulated with broad aims, allowing significant interpretive leeway for teachers, whom were generally trusted to manage their own practices and enjoyed significant autonomy (Mausethagen 2013). Historically, investing in teacher education was a key strategy in guaranteeing education quality (Werler and Sivesind 2007).

In the 1990s, discussions first started regarding quality assessment measures, such as national testing. In 1988, the OECD had published a country review of the Norwegian education system, which questioned how central authorities could monitor and direct such a decentralised education sector, especially when lacking systematic data on education quality and outcomes. While several WPs in the 1990s addressed OECD concerns, and an evaluation system in a decentralised context was considered, there was little concrete action. There existed political controversy surrounding such a system, and lack of support from key political actors, parliamentarians and the main teacher's union (Møller and Skedsmo 2013). However, at the turn of the millennium, the public and political debate on education intensified, contributing to a strong push for political action.

Data and methodology

Our analysis draws on two principal data sources. First, we analysed four key White Papers (WPs) published by the Ministry of Education and Research (MER): WP 30 Culture for Learning (2003-2004), WP 31 Quality in Schools (2007-2008); WP 20 On the Right Track (2012-2013); and WP 21 Eager to Learn (2016-2017). We decided to use WPs as a key data source as these documents form the prime source of political decision-making, providing insight into the official justification behind

the need to revisit existing policies or practices, and the selection and design of policy tools. Similarly, WPs form a key source of information in tracing the policy trajectory and gaining an understanding of the broader policy context. We selected the four above-named WPs after an initial screening of all WPs published by the MER over the past two decades. It was found that these four WPs represented key moments in TBA's trajectory. WP 30 (2003-2004), prepared under right-wing governance, proposed far-going policy changes that would lead to the formal introduction of TBA in Norway. WP 31 (2007-2008), prepared when a red-green coalition returned to office, proposed adjustments and additions to TBA, as did WP 20 (2012-2013). Finally, WP 21 (2016-2017) presents changes undertaken after a right-wing coalition again took office in 2013. Key policy documents published before 2003 (e.g. Moe Commission 1997; MER 2002), as well as legal documents, evaluation reports (e.g. Haug 2003), press releases, and the memoirs of Helge Ole Bergesen (former-State Secretary of Education, 2001-2005) provided background and contextual information for the analysis.

Second, we drew on data from 37 interviews [1] with 40 educational stakeholders and key informants, conducted between September 2017 and February 2018. The interviews served to gain a deeper understanding of the policy process 'behind the scenes', amongst others to identify prominent policy networks and to explore power dynamics and issues related to legitimacy and (de-)politization (Lascoumes and Le Galès 2007). Moreover, the interviews allowed the refinement of regulation theories behind policy tool selection (Maroy 2015), which often remain implicit or unarticulated in formal policy documents.

Purposive and snowball sampling was used to select interview participants. Based on secondary sources and a-priory knowledge of the researchers, a first list of potential interviewees contained participants who were key players in education policy design, formation and/or implementation over the past three decades, experts on recent changes in education policy or influencers in the education debates. To avoid the risk of selection bias, the sample was expanded and refined during the fieldwork process by asking interviewees to identify other potentially relevant participants. Table 1 provides an overview of all participants.

A semi-structured script guided the 60-minute interviews. Amongst other themes, the interviews addressed policy problem(s) intended to be addressed; perceived causes and relevance of these problem(s); promotion of particular policy solutions; rationales and expectations informing policy tool selection; the process of policy design, institutionalisation and policy evolution; administrative traditions and; social values and public opinion. Moreover, emphasis was placed on the role of key

events and actors throughout the policy process, as well as the latter’s motivations, strategies and ideational influences (see Fontdevila 2019). The interviews were audio-recorded and transcribed verbatim.

Table 1. Overview of interview participants

Actors	Number interviewed
Top-level politicians	4
Policy-makers from the Royal Ministry of Education and Research or the Norwegian Directorate for Education and Training	7
Members of government advisory committees	4
Union leaders; members of The Norwegian Association of Local and Regional Authorities (KS)	11
Academics, experts, evaluators	12
Active participants in the public debate on education	2

We combined inductive and deductive approaches during data analysis. Using ATLAS.ti software, three researchers independently performed a first reading of the raw data material, to identify frequent, dominant and significant themes and categories. Based on this first reading, as well as Jessop’s (2010) CPE framework, we developed a codebook, which was used to code all data material structurally during a second reading. This way, we attempted to classify all data material according to the research questions (Saldaña 2009). Based on this structural coding exercise, six main groups of codes were identified: (1) Attributes of the interviewee; (2) Subjective perception of TBA; (3) Policy process: diagnosis; (4) Policy process: policy formation and evolution; (5) Policy process: balance of forces, actors; and (6) Knowledge mobilization. Subsequently, we performed a third, more in-depth reading of all segmented data material, both within and across the identified themes (e.g. MacQueen et al. 1998). During this third reading, we complemented and refined the six macro codes with a set of analytic codes, corresponding to key concepts, actors and mechanisms (Fontdevila 2019).

Findings

In the sections below, we have structured the presentation of our findings according to CPE’s three evolutionary mechanisms (Jessop, 2010). This division allows us to systematically examine our three research questions. By means of the three evolutionary mechanisms, the policy trajectory of TBA is outlined. During the sections ‘variation’ and ‘selection’, the main drivers of the perceived need for policy change and the promotion of TBA as a suitable and desirable instrument are discussed. During the section ‘retention’, the main policy tools by which TBA has been operationalised are presented.

While all reviewed data material has informed the analysis, in this section we have used policy document citations and selected quotes by interviewees [2] to illustrate the ways in which policy changes have been typically described, explained and justified.

Variation: From 'best school in the world' to 'school loser'

Following the results of Norwegian students in the inaugural Programme for International Student Assessment (PISA) 2000, student academic achievement began to dominate the education debate. Norway ended up in 13th place (out of 31) in the international rankings for reading and science, and 17th in mathematics. In reading, the gap between high- and low-performing pupils was greater than in most European countries. The Conservative Minister of Education and Research of the Bondevik II government, Kristin Clemet (2001–2005), who had taken office only six weeks before the PISA findings were published, made a statement reported by all the country's major media outlets: 'This is disappointing, almost like coming home from a Winter Olympics without a single Norwegian medal. And this time we cannot blame the Finns for being doped.' (Dagbladet, December 5, 2001).

Norway's self-image as 'the best school in the world' [3] was crushed, especially by the large disparities in pupils' educational outcomes and the finding that a significant percentage of pupils left compulsory education without basic competencies such as reading (17%). For decades, the comprehensive school model had aimed to provide a school for all, but now it seemed that social and geographic inequalities were reproduced in schools. The subsequent release of other national and international studies reporting similar results (e.g. PIRLS 2001; PISA 2003; TIMSS 2003) contributed to the consensus that Norwegian education had 'a problem'.

When PISA came, I remember that a lot of people said "this is wrong, this must be wrong, this is not...it does not fit with the Norwegian system, with our curricula". When the next PISA study came three years later, everybody said "Yes, we have a problem, we have to do something". The reason why this had changed during these three years was that there was a lot of other information and research, also national research (...), [which] all told us roughly the same. – Interview with Kristin Clemet, Minister of Education and Research (2001–2005), Conservative Party

Public and political education debates sought explanations for the below-expected results. In his memoir *The Battle for the Knowledge School*, Helge Ole Bergesen, who was State Secretary of Education during Clemet's time in office, summarised the problems that afflicted Norwegian schools: 'Norwegian schools seem to have entered a vicious circle where lack of clarity, lack of competence,

low motivation, weak leadership, uncertainty about responsibility and lack of knowledge of results are mutually reinforcing' (Bergesen 2006, 66; authors' translation).

Criticising an era of school governance, it was argued that schools were subject to excessive input regulations, with little awareness of, or responsibility for, results due to the lack of systematic data on pupils' learning outcomes [4]. The PISA results shattered the belief that the desired results would be obtained through the rule of central authorities (Bergesen 2006, 86–87). Discourse emerged characterising Norwegian schools as 'too soft' and 'playful', with limited attention to basic competencies such as reading, writing and numeracy. An evaluation of Curriculum Reform (1997) (Haug 2003) cited another policy problem: education was geared towards the 'average student' and was not sufficiently adapted to the needs of the individual, an ideal of Norwegian schools since the 1930s. This finding was considered unacceptable, especially in an increasingly diverse society. As would later be argued, 'All students are equal, but none of them are alike (...) If we treat everyone alike, we create greater inequality' (MER 2003-2004, 4; authors' translation).

The agenda-setting power of Clemet and her administration was significant. Since the early 1990s, student achievement was a topic of political discussion. However, in the new millennium, learning outcomes dominated debates as the main indicator of education quality. The perceived learning crisis, to which the media contributed significantly, provided a strategic opportunity for the liberal-conservative government coalition to advance a long-desired policy reform:

For us, who had just taken over political leadership of the Ministry of Education and Research, the PISA results were a "flying start". Admittedly, the Conservative Party had complained for a long time about quality problems in Norwegian schools. (...). With the PISA survey, the climate of debate changed abruptly, radically and irrevocably. (Bergesen 2006, 42; authors' translation)

Selection: 'the most research-based reform'

Policy design process

In PISA's aftermath, a variety of contributors proposed policy solutions. Concluding that schools needed modernisation, Kristin Clemet initiated "The school knows best" project to determine ways to improve Norwegian schools (MER 2002). In this project, experts from outside the MER examined international policy research and policy experiences. The project report reviews, amongst other measures, competition and freedom of choice as tools for promoting student achievement and maintaining low government costs. Since the project report was criticised both within and outside the

MER, the members published a second draft that set a milder tone. The following was their overarching conclusion:

We must decentralise responsibility, improve quality control and increase users' empowerment. The school should be controlled from below, not from above, within nationally targeted goals ... We will mobilise for greater creativity and dedication by giving the freedom to take responsibility. (MER 2002, 1; authors' translation)

The Quality Commission, established in December 2001 by the Stoltenberg I government, played an important role in selecting policy instruments to modernise and enhance Norway's education system [5]. Drawing on national and international policy research and expertise (e.g. Granheim and Lundgren 1990; OECD 1988; UNESCO 1990; 1995) and earlier policy proposals (e.g. the Moe Commission 1997), the Commission delivered its first deliberation in June 2002, outlining the framework of a National Quality Assessment System (NQAS). As part of this system, they proposed the establishment of annual national standardised tests and a web portal publishing the results. The publication of the results, so increasing schools' 'visibility' and facilitating comparison and benchmarking, was seen as a significant pressure mechanism for motivating actors and eliciting improvement:

I think many [of us] believed that it is incentive enough in itself that everybody knows that you are doing badly. As to performance-based pay, I cannot remember that we were talking about that. – Interview with member of Quality Commission

After the 'PISA shock', the introduction of national tests was considered necessary by an increasing number of key actors, but controversy persisted regarding the web portal that published school results. While some felt strongly that visibility, comparison, benchmarking and competition would improve education in the long run, this was far from the general view. Nevertheless, the framing of the policy proposal, and the decision-making process, enabled the adoption of both measures in the early 2000s. With regards to the national tests, they had been promoted and framed as an information-gathering tool, allowing for local and individual adaptations following data-based decision-making. This implied that the tests were largely described as a neutral device (cf. Lascoumes and Le Galès 2007), benefiting teachers, pupils, and parents, while the underlying agenda of monitoring, evaluation and control remained masked and under-communicated. Scepticism of teachers about the need for these data and criticisms from test developers about the challenges of the double-purpose test were ignored or dismissed as ideological rather than objective. With regards to the web-portal, it was argued that once data were collected, the results could not be kept secret, pointing to the existing Act on Public Information.

Despite that several actors expressed concern about ‘the possibility that the portal may contribute to ranking’ and how the introduction of tests may guide teaching (MER, 2002-2003), the decision-making process was characterised by a strong sense of urgency. Interviewees contended that government officials’ reaction reflected what Steiner-Khamsi (2003) has described as ‘scandalisation’ in pushing for the contested policy instrument’s adoption. This reaction contributed to the agreement to adopt national tests in reading, writing, numeracy and English and the online web portal during deliberations on the 2003 state budget:

Kristin Clemet used this presentation of the [PISA] test results to create a wave of “this is exceptional, we must act!” So, the decision to implement national tests was never discussed in principle; it was never discussed as a case—it just came as an amendment to the budget proposition [because] the need for action was perceived to be so great. – Interview with former MER policy-maker

The modernisation project’s conclusions, the 2003 Quality Commission’s final proposal, and the evaluation reports of Curriculum Reform 94 (upper-secondary education) and 97 (compulsory education) laid the foundation for the 2004 WP 30 (Culture for Learning) issued by the MER. Reflecting many of the principles dominating educational reforms worldwide, this WP proposed systematic changes in school governance to improve the Norwegian school system. The proposed measures built on five key principles: a) clear national goals; b) knowledge of results; c) clear responsibility placement; d) greater local freedom of action; and e) a good support and guidance system (MER 2003-2004, 25). The NQAS, containing both traditional and new assessment tools, became a key element in the proposed reform.

More concretely, local government and school authorities would be given more decision-making power, including greater local freedom to decide on educational content and working methods, more flexible rules around class and time distribution, and the transfer of negotiations on teachers’ salaries, working hours and employment conditions from state to regional and local authorities. The national tests were presented as an information-gathering tool for the schools, but would also be a key control mechanism for central and local authorities, enabling them to monitor schools’ (and municipalities’) efforts to achieve centrally defined goals. Finally, municipalities would become obliged to establish quality assessment systems to measure and follow up on school performance [6]: teachers, principals and municipal superintendents would be held formally accountable for students’ performance in standardised tests and other quality measures.

Key drivers behind institutional change

Various drivers contributed to policymakers' perceptions of TBA as a necessary and suitable policy instrument. First, interpretations of global economic trends and societal changes paved the way for TBA. In a knowledge economy, a country's educational achievement is considered fundamental to its economic potential and competitiveness, thus competencies such as literacy and numeracy are perceived as essential for individual and nation-state success (MER 2003-2004). Accordingly, school reforms were sought to ensure that basic competencies would be central to educational content, and acquired by all pupils by the end of compulsory education. Moreover, in an increasingly diverse society, individual and local adaptation was considered ever more important (MER 2003-2004). TBA became a key instrument for central authorities to grant greater freedom to local actors while ensuring compliance with government priorities and goals, e.g. regarding pupils' development of basic competencies.

At the same time, the promotion of TBA cannot be seen in isolation from broader changes in public sector governance. TBA is compatible with the NPM policy paradigm, which was formally introduced in Norway in the late 1980s to restructure and modernise public administration. A second wave of NPM reforms subsequently addressed issues of coordination across administrative levels, which were partly caused by public sector fragmentation accompanying the first wave of NPM reforms. NPM allowed government officials to regain oversight over and responsibility for public services, and hence to ensure equity in educational outcomes across social groups. Moreover, both reform waves followed a belief in the need to steer the public sector by means of a performance-oriented culture, in order to raise its efficiency and effectiveness.

This belief was reinforced by an emerging scientific base, to which international organisations such as the OECD contributed, which argued in favour of outcome-based management, accountability and assessment as key measures to modernise and raise the performance of education systems. As such, in the eyes of policymakers, TBA had empirical credibility. Interviewees who were more directly engaged with the adoption and design of reform measures often referred to the OECD reports as an important source of evidence. Nonetheless, the policy document analysis reveals that in addition to international policy research, also national research formed a key source of evidence in reform proposals (e.g. Granheim and Lundgren 1990; Haug 2003).

With the push for evidence-based policymaking, such research documents, as well as advice from external experts, researchers and consultants play an increasingly important role during policy design processes. With regards to the promotion of TBA as a policy solution, expert advice and research documents seem to have served as both an important source of inspiration, as well as justification. In turn, ideological concerns and critique were largely silenced. Nonetheless, despite a tendency towards scientification of policy processes, Norway's political institutions and social-democratic welfare ideologies mediated the selection and retention of TBA tools.

In this light, a final driver in the adoption and institutionalisation of TBA relates to accountability as an 'empty vessel' policy that can be adopted to serve a diverse set of goals (Steiner-Khamsi 2016). Unlike early adopters of TBA, where this policy measure often served to promote market-based reforms (Verger, Parcerisa, and Fontdevila 2019), Norway's TBA was adopted and developed to ensure equity and quality standards in a decentralised education system. The presumed ability of TBA to ensure a basic standard for all, thus equalising disadvantage, contributed to support beyond party political lines. Similarly, adjustments made soon after the adoption of TBA, as explained below, led to broader acceptance of the formerly contested measure, and its eventual institutionalisation in the context of existing regulations and practices.

Retention: 'inclusion as a basic ethical warrant'

WP 30 (Culture for Learning) received unanimous approval by the Norwegian Parliament, laying the foundation for the Knowledge Promotion reform. However, before the Knowledge Promotion could be implemented, there was a shift in government. The liberal-conservative coalition of Bondevik's Second Cabinet was replaced by Stoltenberg's Second Cabinet, a red-green coalition comprised of the Labour Party, the Socialist Left Party, and the Center Party. Some expected the new government coalition to halt the reform, given the significant outcries from teachers, students and even municipalities to postpone it. However, as the new Minister of Education and Research (Øystein Djupedal, member of the Socialist Left Party) explained, they decided to adhere to the principles and the timetable. They shared the view that there was a quality crisis in Norwegian schools and had accepted the logic behind the reform, but also feared the political consequences of abandoning or postponing it.

I do not think the government would have done it anyway. I do not think Jens Stoltenberg [the Prime Minister] would have said “Yes” if I went in and said “the Knowledge Promotion reform is not sufficiently prepared”. He would say that we would look soft on knowledge and skills and that we do not appreciate quality. – Interview with Øystein Djupedal, Minister of Education and Research (2005–2007), Socialist Left Party

One change in the Knowledge Promotion reform implementation related to the administration of national tests and the ‘School Portal’, where results were published. The rushed implementation of the 2004 national tests had contributed to validity problems and destructive evaluation reports. This prompted criticism, even from test proponents, and country-wide boycotts of the test by upper secondary students and the School Student Union.

The School Student Union arranged the boycott against the national test in the spring of 2004 and in January 2005. The basis for the boycott was, on the one hand, opposition to publication of test results and the ideology of competition that this expressed, and secondly, the lack of opportunity to share our views during the planning of the test system. When the authorities neglect students’ opinions and do not allow us to advance our criticisms, warnings and recommendations, it is legitimate to use other forms of influence. (Hølleland 2007, 37; authors’ translation)

The State Secretary to the Minister of Education and Research (2005–2013) admitted that ‘national tests were one of the most difficult issues’ during the negotiations between the three parties on the 2005 political platform. Eventually, they agreed to pause the tests in order to improve their quality and negotiate the conditions of administration. The national tests were reintroduced in 2007, no longer administered during the spring of the 4th, 7th, 10th and 11th grades but instead during the autumn of the 5th, 8th and 9th grades [7]. This decision changed how TBA played out:

So, while the previous tests could be used to test and follow up actual teachers, the new tests [were] intended more to discuss the quality of the school rather than the individual teacher. (...). The most important thing with these tests was to measure school quality on a municipal level, and to provide the municipalities with a good tool to measure quality, and to be able to talk about quality differences between schools, not so much focus on individual teachers but on the school as system. This meant [it] took a lot of pressure away from the actual teachers. – Interview with State Secretary to Minister of Education and Research (2005–2013)

Moreover, the new government coalition saw publication of school results on the ‘School Portal’ as a means of scapegoating teachers, school leaders and local authorities rather than empowering them to improve results. This shaped the decision to no longer publish national test scores at the school level, but only at municipal, county and national levels [8]. At the same time, belief persisted that national test data was necessary for school development, shifting emphasis to the data’s utility in internal discussion, learning and quality improvement. As such, the decision to proceed with national testing

signals acceptance of the policy tool while rejecting its initial use for school comparison and competition.

Nonetheless, while political rhetoric highlighted a belief in the professional responsibility of local actors and educators to use test results to improve schools, control mechanisms have been layered onto one another over time. During eight years of a red-green government coalition (2005–2013), TBA steered towards administrative or bureaucratic forms of accountability, where the consequences of national test scores are determined by national and local authorities rather than by market and competitive pressures. This attempt to weaken the role of external actors, such as the media, in holding schools accountable for results instead caused even greater pressure for local authorities to be liable in educational outcomes.

For example, in WP 31 (Quality in Schools) (MER 2006-2007) a new administrative policy tool requires municipal councils to produce yearly assessments of school academic levels and learning environments and, where necessary, to develop improvement plans. The rationale for intensifying the administrative consequences of school performance was largely, but not solely, based on the perceived detrimental effects of market accountability. According to two policy-makers with key roles in policy formation over the last two decades, the additional pressure on municipalities and schools also reflected parents' failure to hold schools accountable for results:

All these various [assessments], national tests, pupil surveys, value-added indicators require a professional school administration to make use of them. We thought that if parents saw the results of the national tests and of the pupil survey, they might perhaps ... come to the school and say "OK, look here, we have a problem, we are not performing as well as we could or should in the national tests and the important things that they measure; we do not perform as well as we should in terms of the student learning environment", and so on. But my impression is that parents seldom do this, so the system is very dependent on a good, professional and forward-looking school administration. – Interview with MER policy-maker

The Conservative Party returned to office in 2013, this time with the right-wing Progress Party. The new coalition reintroduced the publication of schools' national test scores on the government's 'School Portal'. In addition, they also introduced a new accountability tool as another external device to put pressure on local actors: a quality standard in key areas of education, using the number of students achieving a minimum level of performance as a key indicator (MER 2016-2017). Municipalities and counties averaging below a basic level would receive mandatory support and follow-up. This measure was justified as a means of giving direction and conveying the expectations for the role of local and regional authorities.

In addition, in 2016, value-added indicators for compulsory education were published for the first time at national level. The valuation of a school, municipality or county's contribution to student learning was considered a valuable supplement to information about results from national tests and examinations and a key tool for school development. According to interviewees, municipal superintendents have already welcomed the value-added model as a new accountability tool in their annual talk with the leadership teams of each school.

Finally, adoption of the NQAS and the expectation that municipal superintendents take responsibility for monitoring and following up assessment results afforded opportunities to operationalise and expand the specific governing model. While central policy rhetoric, to a certain extent, expresses trust in the professional responsibility of local and school actors to engage in data-based quality improvement without incentives or sanctions, interpretations differ at the local level. Various municipal superintendents have decided to enact quality assessment systems based on detailed and in advance specified performance indicators, risk assessment, publication of school test scores and performance contracts. In addition, in some large cities, municipal superintendents use merit-based pay, which is part of local salary negotiations, to reward principals who can demonstrate good performance.

Discussion and conclusion

This paper has addressed the following research questions: a) What is the policy trajectory of TBA in Norway? b) What are the main drivers and rationales for Norway's adoption of TBA? c) What are the main policy tools to adopt and develop TBA in Norway? Our analysis portrays that TBA formed a key policy instrument to modernise and raise the performance and equity of the Norwegian education system. TBA replaced a steering tradition based on prescription and intervention, by allowing government officials to steer a highly decentralised education system from a distance, by means of outcome measures, visibility, comparison and accountability.

With regards to the policy trajectory, the analysis reveals that a window of opportunity for major educational reform opened when the arrival of a right-wing government coincided with the publication of the first PISA results. The below-expected results served as an 'external authority' for already existing ideas and policy initiatives (Steiner-Khamsi 2003), and contributed to justify the need to revisit existing policies and practices, and to legitimise the advance of formerly controversial reform

measures. In addition to the ‘scandalisation’ of Norway’s PISA results, to which also the media contributed (see Elstad 2012), the promotion of standardised testing as a neutral device played a key role in the abrupt adoption of national testing in the early 2000s.

Meanwhile, the increasingly dominant policy discourse equalising education quality with academic learning outcomes measured by standardised tests meant that national testing was here to stay. Multiple managerial devices have been introduced to address achievement gaps across different social and cultural groups and, to some degree, the public debate about equity and quality has been rearticulated to performance indicators based on national and international tests. Politicians across the political spectrum have referred to the PISA, as well as national test results, as key measures to judge the education system’s quality. As such, while WP 30 (2003-2004) introduced a transformation in school governance, policy evolution ever since has largely remained characterised by continuity. While adjustments and adaptations have been made during different political cabinets, in particular to downplay competition dynamics and to promote learning, these changes remain characterised by a belief in national testing as a valid measure to assess education quality, promote data-based decision-making and hold key actors accountable.

Beyond the general acceptance of a ‘quality crisis’, which proved a strong catalyst for change, our analysis highlights key drivers and rationales behind Norway’s adoption and retention of TBA. In particular the interpretation of global economic trends and societal changes, TBA’s compatibility with the NPM policy paradigm, and the assumed empirical credibility of TBA form key explanatory factors in this regard. Regarding the latter, the analysis reveals the importance of research documents and expert advice, as both a source of inspiration and legitimisation of reform proposals. Interestingly, while providing policy documents with ‘scientific rationality’, Steiner-Khamsi, Karseth and Baek (2019) show that ‘expert advice’ is referred to in a selective manner in policy documents, highlighting the importance of critically examining the use of research in policymaking.

Nonetheless, despite that national and international research inspired and legitimised reform proposals, Norway’s political institutions and social-democratic welfare ideologies played a key role during the adaptation and evolution of TBA, contributing to its broad acceptance over time. Analysing this semiotically, the fact that global education policy principles such as accountability commonly operate as ‘empty vessels’ is important in this regard. In contrast to early adopters’ use of TBA to promote market-based reforms (Verger, Parcerisa, and Fontdevila 2019), Norway has developed TBA to ensure equity and quality standards in a decentralised education setting.

In this light, our analysis illuminates how TBA has been operationalised by a range of policy tools. Relying on the conception of an actor guided less by self-interest than by social obligations, the Norwegian way of TBA seeks to orient the behaviour of local actors by a combination of external devices that pressure actors, with internal measures designed to mobilise internal feelings of responsibility (Maroy 2015). This finding aligns to the conclusion drawn by Hatch (2013) that the Norwegian accountability system is characterised by a tension between answerability for short-term goals and responsibility for broader purposes. The institutionalisation of TBA furthermore has a clear 'Norwegian touch' in the limited 'hard' consequences attached to school performance, which conflicts with professional values and administrative traditions (Skedsmo 2011). At the same time, despite the political rhetoric implying trust in educators' professional judgement and responsibility, school performance is increasingly monitored, controlled and made visible by administrative mechanisms layered on top of one another over time.

Finally, Norway's TBA is a good illustration of how policy instruments exist 'independent of the decisions that created them' (Kassim and Le Galès 2010, 11), taking forms and generating outcomes that may contradict or extend beyond initial policy goals (Le Galès 2010). In contrast to national policy rhetoric, some municipal superintendents have attempted to re-orient the behaviour of local actors by increasing controls and raising the stakes for school performance. These efforts build on the assumption that many aspects of teaching and learning can be controlled and documented, and that holding school leaders and teachers accountable for students' results will make them more efficient and effective. These local differences highlight the importance of examining the ways in which municipal superintendents act as 'brokers' of predefined goals, in particular when studying school responses to data-use and accountability demands (Prøitz, Mausethagen, and Skedsmo 2019)

To sum up, by tracing the adoption, retention and adaptation of TBA in Norway, this study provides interesting insights into the complex, diverse and hidden drivers of TBA reforms in a social-democratic welfare state. Beyond the study's immediate aims, we identified signs of contradiction and paradox in the enactment of the described reforms. The combined effect of delegating responsibility and decision-making power downward while raising pressure has created tensions over time. As some interviewees argued, although the original aim was to encourage individual and local adaptation and creativity, stricter local authority control and supervision has sometimes constrained teacher autonomy and promoted standardised teaching methods. The ongoing struggle of teachers and school leaders to find meaningful ways of integrating national test data (Mausethagen, Prøitz, and Skedsmo 2018),

moreover, confirms that the challenges associated with assigning a double purpose to a single test have not been resolved (see also Tveit 2018).

Paradoxically, although the Norwegian curriculum guidelines provide a broad framework allowing autonomy of local schools, the scope is narrowed by the municipal use of national standardised tests to hold schools accountable. While policy tools such as national tests may appear neutral, they carry values and meaning, thereby foregrounding certain aspects of teaching and learning while constraining others. For greater understanding of the circumstances and mechanisms contributing to the operation and policy outcomes of TBA modalities, further research is needed to examine how the reforms are interpreted and put into practice on the ground.

Notes

1. There were 34 individual interviews and 3 group interviews with two participants in each.
2. Indirect identification of interviewees is prevented by not referring to a participant's specific role. Interviewees for whom indirect identification is unavoidable, such as previous Ministers, gave explicit consent to be quoted and named.
3. This conclusion, drawn in 1975 by Bjartmar Gjerde, Minister of Church and Education Affairs from 1973 to 1976, reflected a widely shared belief.
4. That is, while much was known concerning what was going into schools in terms of resources, teachers, pupils, buildings, computer equipment etc., barely anything, except final exam scores, was known about what was coming out.
5. Upon establishment, the Commission had been mandated to 'evaluate the content, quality, and organisation of basic education'. Under the Bondevik II government, 8 members were added to the Commission, while the mandate was extended and now included the request to develop a quality assessment system for basic education.
6. As a result of NPM reforms introduced during the late 1980s, many municipalities had already established quality assessment systems.
7. This meant that tests were not reintroduced at the upper-secondary level but only in compulsory education.
8. Nevertheless, results could still be requested by anyone under the Freedom of Public Information Act.

References

- Ball, S. J. 1990. *Politics and Policy-making in Education: Explorations in Policy Sociology*. London: Routledge.
- Ball, S. J. 1998. "Big Policies/Small World: An Introduction to International Perspectives in Education Policy." *Comparative Education* 34, no. 2: 119–130. doi:10.1080/03050069828225
- Ball, S. J., C. Junemann, and D. Santori. 2017. *Edu-net: Globalisation and Education Policy Mobility*. London: Routledge.
- Bergesen, H. O. 2006. *Kampen om kunnskapsskolen [The Battle for the Knowledge School]*. Oslo: Universitetsforlaget.
- Blossing, U., G. Imsen, and L. Moos. 2014. *The Nordic model. A 'School For All' Encounters Neo Liberal Policy, Policy Implications of Research in Education (Vol. 1)*. Dordrecht: Springer Series.
- Christensen, T., and P. Læg Reid, 2011. "Democracy and Administrative Policy: Contrasting Elements of New Public Management (NPM) and post-NPM." *European Political Science Review* 3, no. 1: 125–146. doi:10.1017/S1755773910000299
- Dale, R. 1999. "Specifying Globalisation Effects on National Policy: A Focus on the Mechanisms." *Journal of Education Policy* 14, no. 1: 1–17. doi:10.1080/026809399286468
- Elstad, E. 2012. "PISA Debates and Blame Management among the Norwegian Educational Authorities: Press Coverage and Debate Intensity in the Newspapers." *Problems of Education in the 21st Century* 48: 10–22.
- Esping-Andersen, G. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.
- Fontdevila, C. (2019). *Researching the adoption of School Autonomy with Accountability reforms: A methodological note on country case studies*. REFORMED Methodological Papers No. 1. doi: 10.5281/zenodo.3557302
- Granheim, and Lundgren, S. 1990. *Goal management and evaluation in the Norwegian school. Final report EMIL project*.
- Gunter, H., D. Hall, R. Serpieri, and E. Grimaldi. 2016. *New Public Management and the Reform of Education: European Lessons for Policy and Practice*. London: Routledge.
- Hall, D., E. Grimaldi, H. Gunter, J. Møller, R. Serpieri, and G. Skedsmo. 2015. "Educational Reform and Modernisation in Europe: The Role of National Contexts in Mediating the New Public Management." *European Educational Research Journal* 14, no. 6: 487–507. doi:10.1177/1474904115615357
- Hamilton, L. S., B. M. Stecher, and S. P. Klein. 2002. *Making Sense of TBA in Education*. Santa Monica, CA: Rand Corporation.

- Hatch, T. 2013. "Beneath the surface of accountability: Answerability, responsibility and capacity building in recent educational reforms in Norway." *Journal of Educational Change* 14, no. 1: 113-138. doi:10.1007/s10833-012-9206-1
- Haug, P. 2003. *Evaluering av Reform 97: Sluttrapport frå styret for Program for evaluering av Reform 97* [Evaluation of Reform 97: Final report from the board of the Program for Evaluation of Reform 97]. Oslo: Norges forskningsråd.
- Haugen, C. R. 2019. "Teachers' experiences of school choice from 'marginalised' and 'privileged' public schools in Oslo." *Journal of Education Policy*. doi:10.1080/02680939.2019.1587519
- Hølleland, H. 2007. "Nasjonale prøver og kvalitetsutvikling i skolen" [National tests and quality development in the school]. In *Elevvurdering i skolen* [Student assessment in education], edited by S. Tveit, 29-43. Oslo: Universitetsforlaget.
- Hovdenak, S. S., and J. H. Stray. 2015. *Hva skjer med skolen? En kunnskapssosiologisk analyse av norsk utdanningspolitikk fra 1990-tallet og frem til i dag*. [What happens to the schools? A sociological analysis of Norwegian education policy from the 1990s until today]. Bergen: Fagbokforlaget.
- Jessop, B. 2010. "Cultural Political Economy and Critical Policy Studies." *Critical Policy Studies* 3, nos. 3-4: 336–356. doi:10.1080/19460171003619741
- Kassim, H., and P. Le Galès. 2010. "Exploring Governance in a Multi-Level Polity: A Policy Instruments Approach." *West European Politics* 33, no. 1: 1–21. doi:10.1080/01402380903354031
- Langfeldt, G., E. Elstad, and S. T. Hopmann. 2008. *Ansvarlighet i skolen* [Accountability in schools]. Oslo: Cappelen Forlag
- Lascoumes, P., and P. Le Galès. 2007. "Introduction: Understanding Public Policy through Its Instruments—From the Nature of Instruments to the Sociology of Public Policy Instrumentation." *Governance: An International Journal of Policy, Administration, and Institutions* 20, no. 1: 1–21. doi:10.1111/j.1468-0491.2007.00342.x
- Le Galès, P. (2010). "Policy Instruments and Governance." In *The SAGE Handbook of Governance*, edited by Mark Bevir, 142–160. SAGE Publications.
- Lenschow, A., D. Liefferink, and S. Veenman. 2005. "When the Birds Sing: A Framework for Analysing Domestic Factors behind Policy Convergence." *Journal of European Public Policy* 12, no. 5: 797–816. doi:10.1080/13501760500161373
- Maroy, C. 2015. "Comparing Accountability Policy Tools and Rationales: Various Ways, Various Effects?" In *Governing Educational Spaces: Knowledge, Teaching, and Learning in Transition*, edited by H. G. Kotthoff and L. Klerides. Rotterdam: Sense Publication.

- Mausethagen, S. 2013. "A Research Review of the Impact of Accountability Policies on Teachers' Workplace Relations." *Educational Research Review* 9: 16–33. doi:10.1016/j.edurev.2012.12.001
- Mausethagen, S., T. S. Prøitz, and G. Skedsmo. 2018. "Teachers' Use of Knowledge Sources in 'Result Meetings': Thin Data and Thick Data Use." *Teachers and Teaching: Theory and Practice* 43, no.6: 37–49. doi:10.1080/13540602.2017.1379986
- MacQueen, K. M., E. McLellan, K. Kay, and B. Milstein, 1998. "Codebook development for team based qualitative analysis." *Cultural Anthropology Methods*, 10 no. 2: 31–36. doi: 10.1177/1525822X980100020301
- Ministry of Education and Research (MER). 2002. *Skolen vet best! [The School Knows Best!]*. Oslo.
- Ministry of Education and Research (MER). 2002-2003. Appendix no. 3. Retrieved from: <https://www.regjeringen.no/no/dokumenter/stprp-nr-1-tillegg-nr-3-2002-2003-/id435850/?ch=2>
- Ministry of Education and Research (MER). 2003-2004. *Kultur for læring [Culture for Learning]*. White Paper No. 30.
- Ministry of Education and Research (MER). 2007-2008. *Kvalitet i skolen [Educational Quality]*. White Paper No. 31.
- Ministry of Education and Research (MER). 2012-2013. *På rett vei [On the Right Track]*. White Paper No. 20.
- Ministry of Education and Research (MER). 2016-2017. *Lærelyst [Eager to Learn]*. White Paper No. 21.
- Møller, J., and G. Skedsmo. 2013. "Modernising education: New Public Management reform in the Norwegian education system." *Journal of Educational Administration and History* 45 no. 4: 336–353. doi:10.1080/00220620.2013.822353
- OECD. 1988-89. *Ekspertvurdering fra OECD. OECD-vurdering av norsk utdanningspolitikk [Expert review from the OECD: OECD review of Norwegian education policy]*.
- Prøitz, T., S. Mausethagen, and G. Skedsmo. 2019. "District administrators' governing styles in the enactment of data-use practices." *International Journal of Leadership in Education*: 1-22. doi:10.1080/13603124.2018.1562097
- Quality Commission. 2002. *Førsteklasses fra første klasse. Forslag til rammeverk for et nasjonal kvalitetsvurderingssystem av norsk grunnsopplæring [Proposal for framework for the national system of quality assessment of the Norwegian comprehensive education]*. Green Paper No. 10.
- Quality Commission. 2003. *I første rekke. Forsterket kvalitet i en grunnsopplæring for alle [Improving quality in comprehensive education for all students]*. Green Paper No. 16.

- Sahlberg, P. 2016. "The Global Education Reform Movements and its Impact on Schooling." In *The Handbook of Global Education Policy*, edited by K. Mundy, A. Green, B. Lingard, and A. Verger, 128–144. Oxford: Wiley Blackwell.
- Saldaña, J. 2009. *The Coding Manual for Qualitative Researchers*. London: Sage.
- Steiner-Khamsi, G. 2003. "The Politics of League Tables." *Journal of Social Science Education*, 2003 -1. <http://www.jsse.org/index.php/jsse/article/view/470>
- Steiner-Khamsi, G., B. Karseth, and C. Baek. 2019. "From science to politics: commissioned reports and their political translation into White Papers." *Journal of Education Policy*. doi:10.1080/02680939.2019.1656289.
- Skedsmo, Guri. 2009. "School governing in transition? Perspectives, purposes and perceptions of evaluation policy." PhD diss., University of Oslo.
- Skedsmo, G. 2011. "Formulation and realisation of evaluation policy: Inconsistencies and problematic issues." *Educational Assessment, Evaluation and Accountability* 23 (1): 5–20. doi:10.1007/s11092-010-9110-2
- Tveit, S. 2018. "Ambitious and ambiguous: shifting purposes of national testing in the legitimation of assessment policies in Norway and Sweden (2000–2017)." *Assessment in Education: Principles, Policy & Practice* 25 no. 3: 327-350, doi:10.1080/0969594X.2017.1421522
- UNESCO. 1990. *Planning for Quality*. Paris: UNESCO.
- UNESCO. 1995. *Schools under scrutiny*. Paris: UNESCO.
- Verger, A., C. Fontdevila, and A. Zancajo. 2016. *The Privatisation of Education. A Political Economy of Global Education Reform*. New York: Teachers College Press.
- Verger, A., L. Parcerisa, and C. Fontdevila. 2019. "The Growth and Spread of National Assessments and Test-based Accountabilities: A Political Sociology of Global Education Reforms." *Educational Review* 71, no. 1: 5–30. doi:10.1080/00131911.2019.1522045
- Werler, T., and K. Sivesind. 2007. "Norway." In *The Education System of Europe*, edited by H. Döbert, W. Hörner, B. von Kopp, and W. Mitter, 573-589. Dordrecht: Kluwer.

Publication B: Camphuijsen, M. K., & Levatino, A. (2020, under review). Schools in the media: framing national standardized testing in the Norwegian press, 2004-2018. *Discourse: Studies in the Cultural Politics of Education*.

CHAPTER 3

Schools in the Media: Framing National Standardized Testing in the Norwegian Press, 2004-2018

Introduction

In recent decades, media outlets have been increasingly active in reporting on standardized testing. A growing body of research has examined how the media cover standardized tests, as well as with what implications, focusing both on international large-scale assessments (e.g. Baroutsis & Lingard, 2017; Dixon et al., 2013; Elstad, 2012; Stack, 2006; Waldow, Takayama, & Sung, 2014) and on national standardized tests (e.g. Baroutsis, 2016; Mockler, 2013; 2020; Shine & O'Donoghue, 2013; Shine, 2015; Yemini & Gordon, 2015). While these studies cover diverse educational settings, non-Anglophone contexts remain under-researched, in particular with regards to media coverage of national standardized testing. In addition, many studies have limited their analyses to a small sample of newspaper articles and/or a selection of national newspapers, while the regional and local press have been more rarely considered.

Nevertheless, in some countries, the local press is particularly important in providing local communities with information, which is viewed as more relevant and 'closer' to home (House of Lords, 2008). This is the case in Norway, which has a strong local newspaper tradition (Østbye, 2019). In 2017, regional and local newspapers in Norway made a total of 1.69 million distributions⁴ (Norwegian Media Authority, 2018), i.e., reaching 42% of the population over the age of 20 (Statistics Norway, 2019). In a context such as the Norwegian one, with a very dispersed population across an extensive geographical area, an exclusive focus on national newspapers signifies the neglect of an important part of the whole media ecosystem. Moreover, the impact of regional and local press on readers may be greater because the presented content is closer to home.

The paucity of studies analyzing longitudinal samples of newspaper articles and the under-consideration of the regional/local press limit the understanding of how media coverage evolves over

⁴ Of these distributions, 1.4 million included paid-for subscriptions, while nearly 300,000 distributions comprised free newspapers. For seven free newspapers, no data can be found on the number of distributions (Norwegian Media Authority, 2018).

time and varies across geographical localities. To address these gaps, we draw upon a corpus of 3,046 articles on national tests, published in Norwegian regional and local newspapers between 2004 and 2018. We aim to (a) identify the most recurrent discursive frames used by the media when informing readers about national standardized tests, as well as their main features, and (b) examine how pervasive certain frames are and how they evolve over time and are distributed across geographical localities. In so doing, we rely on framing theory (Entman, 1993), which forms a useful conceptual framework for understanding how one issue can be presented and framed in different ways, thereby communicating highly distinct causal and normative beliefs and realities (Chong & Druckman, 2007). It is important to gain a deeper understanding of how educational issues, such as standardized testing, are framed, since previous research has highlighted the impact that framing can have on public understanding and opinions (Chong & Druckman, 2007; 2010), as well as on public policy processes (Baroutsis & Lingard, 2017).

National testing in Norway

In Norway, the municipalities are responsible for the 10 years of compulsory education, following the division of school governance responsibilities between central, county and municipal authorities. Introduced in 2004, national standardized tests constitute a central steering device to monitor whether schools and municipalities achieve central learning objectives (Skedsmo, 2011). In addition, they are meant to provide useful information to teachers, school leaders, parents, students and municipal authorities, thereby fostering school improvement efforts. Alongside the adoption of the tests, the government website the ‘School Portal’ was launched in 2004, and aggregated test results are published here. While the tests are one element of the country’s quality assessment system (Skedsmo, 2011), they have become the prime measure to hold schools and municipalities accountable for student learning (Skedsmo & Mausestagen, 2016).

Even though national tests had long been considered controversial in Norway (Tveit, 2014), the PISA shock of 2001 contributed to a relatively broad consensus among political parties and stakeholders about the need to introduce large-scale assessments in the country. Nonetheless, the speed with which the tests were implemented, the lack of consultation of key stakeholders regarding the tests’ design and administration, and the decision that test scores would be published, were heavily criticized (Camphuijsen, Møller & Skedsmo, 2020). The first two test rounds were met by student boycotts and

destructive evaluation reports (Lie et al., 2005), leading the new government coalition to pause the administration of the tests in 2006, in order to improve their quality. In 2007, the tests in reading, numeracy and English were reintroduced. Since then, the annual tests have been taken at the beginning of the fifth, eighth and ninth grade. All students are required to participate, but schools can exempt students with the right to Norwegian language instruction or special needs education. From 2007 until 2014, under red-green governance, only municipal, county and national test results were published on the ‘School Portal’, while school results remained hidden⁵. Since 2014, when a Conservative government coalition returned to office, school results have again been published.

Examining media coverage on standardized testing through frame analysis

Drawing on the seminal work of the sociologist Erving Goffman (1974), framing theory was adapted and developed by scholars such as Iyengar (1990; 1991) and Entman (1993) as a tool to examine how the media organize and present information on particular issues. According to Entman (1993, p. 51), to frame is:

...to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.

Frames thus promote a particular version of reality, and do so by relying on different framing devices, such as tone, placement, repetition, keywords or culturally familiar symbols or images (Entman, 1993), as well as catch phrases (Pan & Kosicki, 1993) and metaphors (Gamson & Modigliani, 1989). As important as the selection and salience of particular aspects of a described reality, can be the exclusion of other aspects (Entman, 1993). By examining the four framing ‘functions’, i.e., problem definition, causal interpretation, moral evaluation and treatment recommendation (Entman, 1993), as well as the framing devices present in a text, a deep understanding can be gained of the ‘interpretative lenses’ through which media texts can guide readers to notice, understand, evaluate and act upon a particular problem in a particular way (Entman, 1991; Chong & Druckman, 2007; Scheufele & Tewksbury, 2007).

⁵ Even though school results remained hidden on the ‘School Portal’, anyone interested could request access to these data, following the Freedom of Information Act.

In recent decades, a range of studies have relied on framing analysis to examine how education, schools and educators are represented and framed in media texts (e.g. Baroutsis, 2016; Baroutsis & Lingard, 2017; Goldstein, 2010; Ulmer, 2014). In this way, it has been revealed how media reports can create and perpetuate a discourse of professional inadequacy (Ulmer, 2014), as well as visually and textually frame and reinforce particular reforms as the sole solution to the crisis in public education (Goldstein, 2010). Other studies have identified how different frames can be present in media coverage on one educational issue (Baroutsis, 2016; Baroutsis & Lingard, 2017). These studies form important contributions to the understanding of how the media frame educational issues. Non-Anglophone contexts, however, remain under-researched.

Data and methodology

Our analysis has been conducted on a corpus of articles published in Norwegian regional and local newspapers from 2004 (the year the tests were introduced) to 2018.

To obtain a complete list of all newspapers, we used Høst's media report (Høst, 2018) and data derived from the Norwegian Media Authority, the National Association of Local Newspapers and the Norwegian Media Businesses' Association. We selected the newspapers according to their geographical distribution. Regional newspapers are those covering one or more counties, whereas local newspapers are those covering districts, cities or one or more municipalities. To examine whether the identified newspapers had reported on national testing during the 15 years being researched, we relied on the digital archive of each newspaper. This approach enabled us to obtain a more complete corpus than if we had used a media search engine such as Retriever, whose coverage of local newspapers is not comprehensive. We used the term 'national test(s)' to search all digital archives. A first scan of all retrieved articles was then performed to select only those articles with a clear main focus on national testing, i.e., (a) articles that reported news about national testing; (b) articles that expressed opinions/debates on national testing; (c) articles that reported national test scores⁶. Of all 202 regional and local newspapers, 155 (77%) had published articles focusing on national testing during the selected publication time frame. In total, 3,046 articles were retrieved after this first scan.

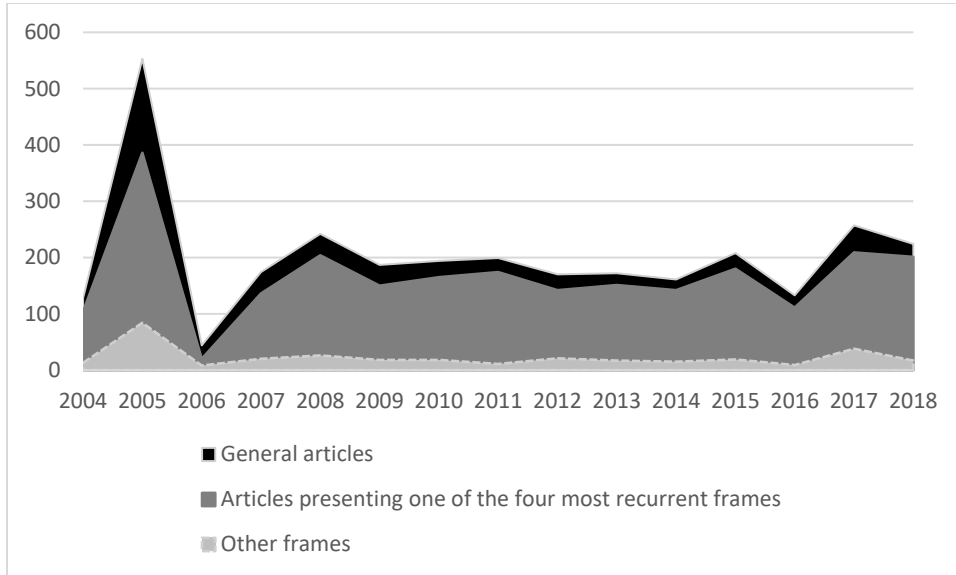
⁶ We included all types of contributions, so the corpus includes articles written by journalists, editorials and opinion and commentary articles.

Informed by framing theory, we operationalized each of the four framing functions (Entman, 1993), namely (1) problem definition, (2) causal interpretation, (3) moral evaluation, (4) treatment recommendation. Based on this operationalization, we then coded all the articles using a codebook, which was the product of a dynamic, nonlinear approach. By coding for framing functions, rather than frames, one is able to identify distinct frames in a more valid and reliable manner (Matthes & Kohring, 2008). In addition to the four framing functions, the continuous iterations with the analyzed texts led us to inductively generate and add new categories to the initial ones (e.g. codes were used to identify score reporting, ranking, naming, praising, blaming, etc.). Parallel to the coding process, a longitudinal dataset was created, which included information on every article of the corpus, i.e., article identifier, title, name of the newspaper where it was published, geographical distribution and date of publication. In this database, all information about the codes used in every single article was inserted. Once we had coded all the articles, the longitudinal dataset enabled us to quantify the recurrence and coexistence of the codes. By means of different combinations of codes, we were able to identify different frames. We then relied on qualitative content analysis (McNamara, 2005) to gain a deeper understanding of the frames identified. Finally, we analyzed the dataset to disclose trends across place and time.

Framing national standardized testing in the Norwegian press

Figure 1 provides a longitudinal overview of the analyzed corpus and shows the evolution of press coverage of national testing over time. The figure shows a peak in press coverage in 2005, one year after the tests were introduced, and a sharp decline in 2006, when the administration of the tests was paused. Ever since the reintroduction of the tests in 2007, press coverage has remained relatively stable.

Figure 1. Number of articles in the analyzed corpus, 2004-2018



Four recurrent frames could be clearly identified in our analysis. These were present in 2,475 newspaper articles (81.25% of the corpus). The remaining articles were articles where other, non-recurrent frames were present (334 in total), or articles that merely reported general information about the tests, such as practical information about the tests' administration (237 in total). In the presentation of the findings that follows, we focus on these four most recurrent frames.

Four most recurrent frames: evolution over time and geographical differences

We named the four dominant frames 'performance', 'transparency and empowerment', 'misinterpretation and misuse' and 'criticism'. Table 1 outlines each frame according to the four framing functions (Entman, 1993).

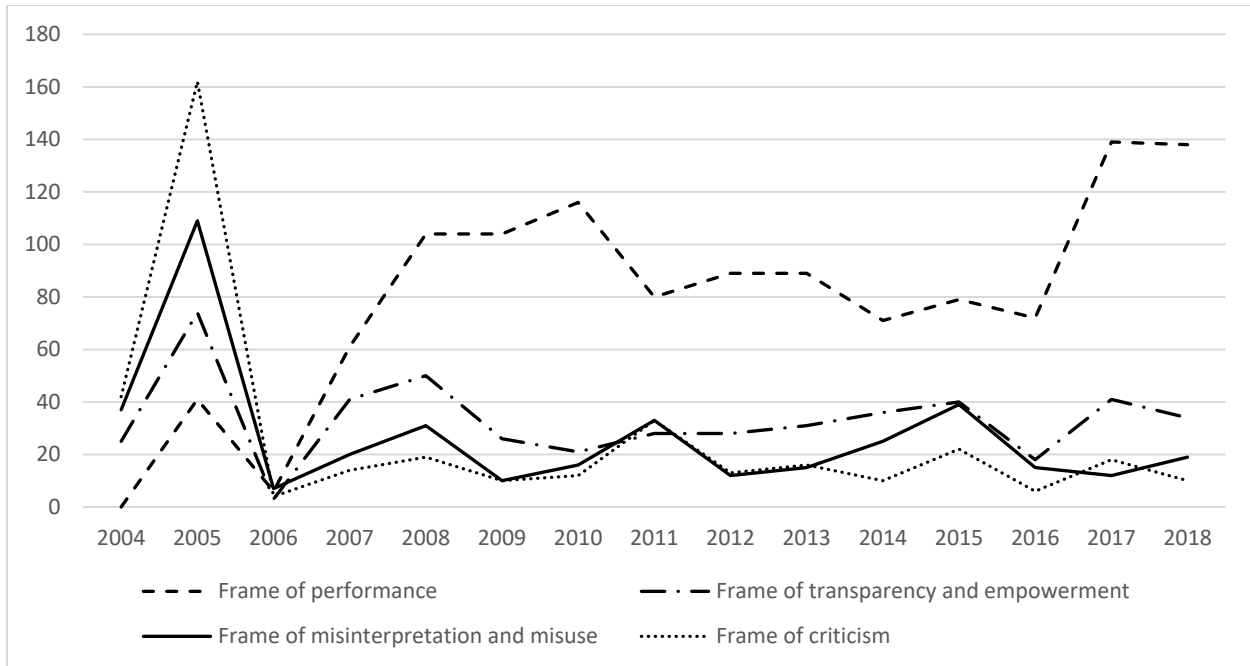
Table 1. Overview of the four most recurrent frames

	Problem definition	Causal interpretation	Moral evaluation	Treatment recommendation
Frame of performance	<p>Everyone needs basic skills to be able to succeed in life. Failure to acquire these skills forms both an individual and a societal problem.</p> <p>National tests measure basic skills. Low test performance reflects failure to acquire them and is alarming.</p>	<p>National test performance reflects the effort put by key actors in the acquisition of basic skills.</p> <p>High-performers found the right formula to acquire basic skills and succeed in the test.</p>	<p>National test scores are valid and relevant to measure, compare and judge the acquisition of basic skills.</p> <p>As test performance reflects effort, actors can be praised or blamed for their performance.</p>	<p>Identification of recipes for success that low-performers can/should learn from high-performers.</p> <p>Those responsible for poor performance should urgently address this problem and take measures to assure the acquisition of basic skills.</p>
Frame of transparency and empowerment	<p>There are large inequalities in student learning (between students, schools, and localities).</p> <p>It is crucial to know more about these inequalities to address them.</p>	<p>National tests and transparency in terms of learning outcomes guarantee that key actors (policymakers, stakeholders and society as a whole) become aware of these inequalities, and are empowered to address them.</p>	<p>National tests are valid, reliable and useful instruments to uncover students' learning difficulties and adapt education to individual and local needs.</p> <p>Transparency in terms of learning outcomes is both a democratic right, and a prerequisite for school improvement.</p>	<p>Safeguarding of the national testing system.</p> <p>Guarantee of transparency in terms of learning outcomes.</p>
Frame of misinterpretation and misuse	<p>National test results are often misinterpreted and/or misused.</p>	<p>National test results are often interpreted as prime indicators of education quality, but they only measure a set of learning goals.</p> <p>Statements that are not statistically sound are often made.</p> <p>Because of the publication of results,</p>	<p>National tests can be useful educational tools, but serious problems emerge when test results are misinterpreted or badly used.</p> <p>There is a wrong way and correct way to use the test.</p>	<p>Correct interpretation and use of test results for pedagogical and organizational purposes.</p> <p>Abandonment of publication of results.</p>

		national test results have often become a goal in itself, generating pressure, competition and undesired practices.		
Frame of criticism	National test system is disruptive and has a negative impact on education quality and equity.	The national test system was demanded and developed by politicians, rather than school professionals, and fails to be a valuable pedagogical and organizational tool for schools. The national test system causes the adoption of harmful practices and has several negative side-effects.	Delivered national test data are not valid, not reliable and/or not useful. The national test system is detrimental to education quality and equity. No right use of national test data is possible.	Withdrawal of the national test system. Promotion of assessment measures considered suitable by school professionals and reliance on educators' professional judgement, experience and knowledge.

Often (in 113 cases), the same newspaper contains articles that offer different, sometimes competing, frames. That is, even newspapers that regularly report results and rankings, publish articles that harshly criticize this practice, as well as articles that present strong critiques of the validity and reliability of the test results. Nonetheless, the four frames identified were not equally present in our corpus, and their dominance varied over time and across geographical localities. Figure 2 displays the evolution of the four frames over the analyzed 15-year time span.

Figure 2. Number of articles presenting the four most recurrent frames, 2004-2018

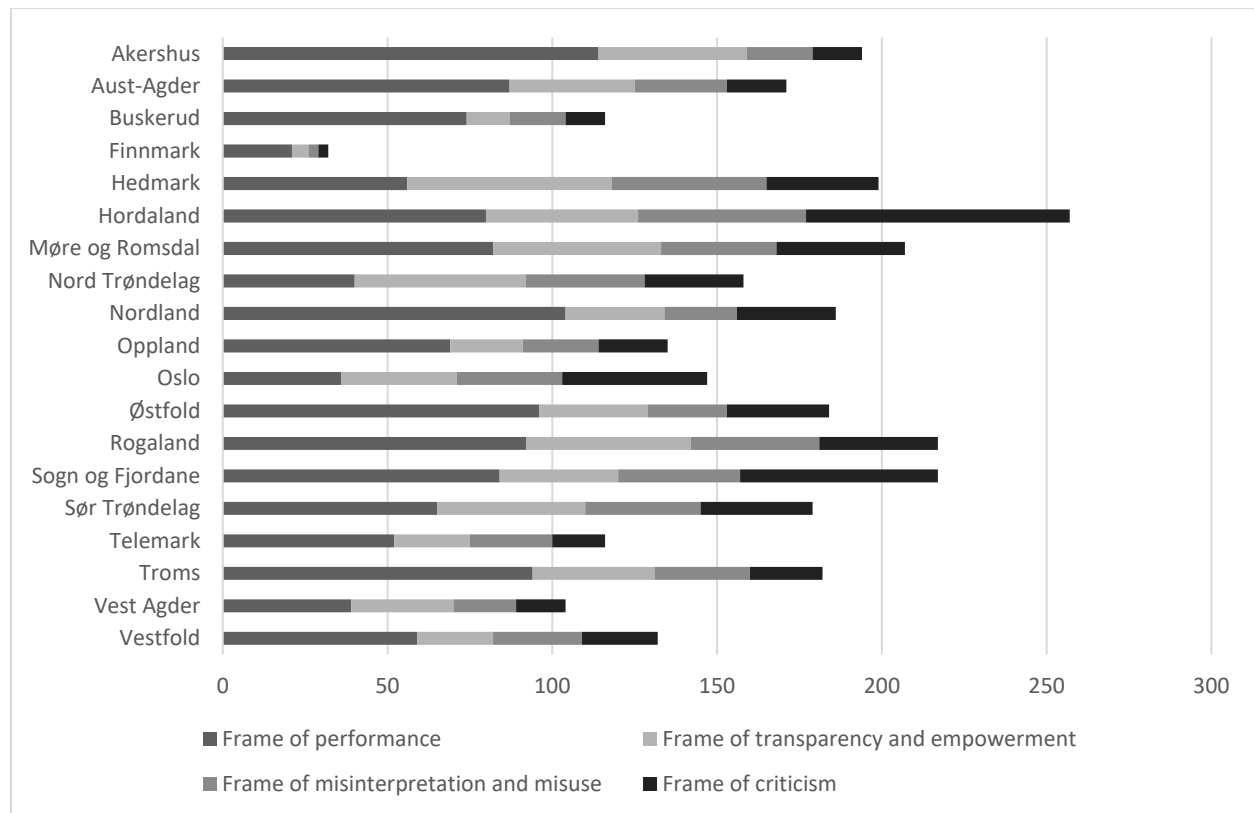


As depicted in the figure, the frames of ‘criticism’ and ‘misinterpretation and misuse’ were particularly pervasive during the years of the first two test rounds. Their dominance, however, declined upon the reintroduction of the tests in 2007. Ever since, the frame of ‘performance’ has become the most dominant frame in the articles analyzed, followed by the frame of ‘transparency and empowerment’.

In addition to differences in time, the analysis identified differences in how pervasive each frame is across geographical localities. Figure 3 shows the recurrence of each of the four frames per county⁷.

⁷ As of January 2020, the former 19 counties were merged into 11 counties. However, as the former county distribution was in place during the analyzed time span, and the official data on the geographical distribution of newspapers relies on this county distribution, we present the data for the 19 former counties.

Figure 3. Geographical distribution of articles presenting the four most recurrent frames (county-level)



As can be derived from the figure, geographical differences can be observed with regard to the absolute number of articles published on national testing in each county, as well as with regard to the dominance of each of the four frames. In 15 counties, the frame of ‘performance’ formed the most dominant frame, while in three other counties (Hedmark, Nord-Trøndelag and Oslo), another frame was found slightly more often. Hordaland forms the sole example of a county where two frames, namely ‘performance’ and ‘criticism’, were found to be present in the same number of articles. Even though in Oslo the frame of ‘criticism’ was slightly more frequent, the county represents a unique case of a near balance in the dominance of all four frames.

In the next subsections, we present each discursive frame in more depth, using salient quotations to epitomize and explain the frames⁸.

The frame of 'performance'

The frame of 'performance' was the most frequently found frame in the Norwegian regional and local press in the time span analyzed (present in 1,190 articles). As shown in Table 1, the frame of 'performance' communicates the idea that national tests measure the acquisition of basic skills that are crucial for individual success in life and for societal progress. Consequently, low test performance should 'ring the alarm bells' (Solvang, 2017). The frame drives the message that test performance, to an important extent, reflects the effort put in by key actors to secure acquisition of basic skills (most notably students, educators and municipal authorities), and that high-performers have 'found the formula' to acquire basic skills (Hovland, 2018). This message reflects an implicit evaluation of national test scores as valid indicators to measure, compare and judge students' acquisition of basic skills. Furthermore, as performance in the national tests corresponds to the effort put in by key actors, those who obtain low scores can be criticized or even blamed. Finally, as a treatment recommendation, the frame conveys the message that those responsible for the low results urgently have to take measures to raise them and that they can learn from high-performers, who have found measures that 'work' (Solvang, 2015). Below, we provide examples of how the frame of 'performance' disseminates and reinforces the above-outlined messages.

In total, 1,117 articles (93.87% of the 1,190 presenting this frame) use test scores to make any form of comparison between the performance of different entities (e.g., school to school, school to national average, municipality to county, etc.), or over time. Even though school results were not published on the 'School Portal' between 2007 and 2014, this did not hamper the construction of school comparisons and rankings in the regional and local press during this period⁹.

⁸ Considering the size of the corpus, it was not possible to represent all of the analyzed articles. We therefore selected quotations that represent typical examples of commonly expressed statements. All quotations have been translated from Norwegian to English by the authors.

⁹ Our analysis shows that, even though school results were not published on the 'School Portal', during this period, 619 articles of our corpus were presenting school comparisons and/or rankings. It was found that in 18 cases, school rankings were based on results presented in national newspapers, in 9 cases school results were provided by county/municipal authorities, and in 3 cases local newspapers requested school data following the Freedom of Information Act. In all other cases, the source of the school data was not provided.

Test scores are occasionally presented in ways that encourage the reader to think of the national testing system as a competition of some sort. For example, in 167 articles the frame of ‘performance’ is constructed around the conceptual metaphor (cf. Lakoff & Johnson, 2003) ‘national test performance is a competitive sport’. That is, terms and language taken from sports journalism are used to report on school test results. Examples include: ‘Students break their own personal record’ (Lofotposten, 2013) and ‘*Municipality X* is beaten at the finish line by *Municipality X* and *Municipality X*’ (Frimand, 2008). As argued by Schmitt (2005, p. 336) metaphors make ‘perception more automatic and ease the energy required to understand.’ However, the use of this conceptual metaphor not only enhances the power of the message and catches the reader’s attention, but also conveys a range of implicit messages. For instance, sport is quintessentially meritocratic and sport competition is healthy and aimed at self-overcoming. Accordingly, the best performers can/should be praised.

In this regard, in the frame of ‘performance’, a judgment is often assigned to the comparative position of one or more entities, blaming those that score below the desired level, while praising those that perform beyond expectations. Of the 1,172 articles within this frame that name schools, municipalities and/or counties (i.e., 98.49%), 235 (20.05%) articles include blame and 570 (48.63%) articles include praise. Examples of blaming include statements such as ‘School losers in *Municipality X* and *Municipality X*’ (Hermstad, 2008) and ‘Worst in the country at 2 + 2’ (Eriksen, 2011). In contrast, examples of praising include statements such as ‘*School A* tops the list of Norway’s best schools’ (Arbeidets Rett, 2018) and ‘No one is better at counting than this group’ (Findahl, 2017). This result confirms the observation by Elstad (2009) that Norwegian newspaper articles often focus on the worst and/or best-performing schools. It also seems to correspond to the tendency towards sensationalization, which previous research has pointed out as a fundamental feature of media reporting (Baroutsis, 2017). However, in contrast to what has been reported by previous research on a negative press bias, i.e., a bias towards reporting negative facts (Dixon et al., 2013; Elstad, 2009; 2012), we found praising a more common practice than blaming.

The frame of ‘transparency and empowerment’

The frame of ‘transparency and empowerment’ was the second most frequently found frame, present in 495 articles. The frame problematizes that there exist significant inequalities in student learning in Norwegian education. Before the adoption of the testing system, as well as during periods that test

scores were not published at the school level, these inequalities have largely gone unnoticed, and have therefore remained unaddressed¹⁰. National tests provide knowledge that is presented as essential to uncover these inequalities and empower key actors (e.g. policymakers, educators and society as a whole) to address them, thereby making ‘winners of those who might otherwise become losers’ (Sørensen, 2004).

National tests are thus evaluated as valid, reliable and useful instruments to uncover students’ learning difficulties, adapt education to individual and local needs, and hold teachers and other actors accountable for low scores. What follows from this positive evaluation of testing and transparency to address social inequalities, is the recommendation to safeguard the testing system and to guarantee transparency in terms of learning outcomes. In turn, arguments against and efforts to discontinue testing and/or the publication of results are harshly criticized:

It is not the tests that create the inequalities - but it is the tests that reveal them, thus making it possible to direct efforts to help the most struggling schools and students. (...). The differences do not disappear even if you hide them away. They only become invisible and continue to destroy the future of new generations. (Skilleås, 2005)

One important assumption behind the requirement for transparency is that it can serve as an important pressure mechanism for schools and municipalities to reflect on their work, make changes and improve results. Hiding the results, in contrast, would imply that quality improvement measures ‘lie untouched in archives, drawers and folders in public offices and with teachers’ (Borgen, 2009). An important second reason behind the promotion of transparency is that it is a democratic right and it facilitates public engagement and deliberation, allowing school administrators, parents, local politicians, journalists and the general public to follow and engage with the efforts of municipalities, schools and teachers to ensure education quality and equity. In 114 of the 495 articles within this frame (i.e., 23.03%), the publication of results is promoted as a key condition to assure transparency and for the tests to be of optimal value.

¹⁰ In making this claim, no mention is made of the fact that, as highlighted in footnote 6, some media and municipalities reported school results during the years that they were not published on the ‘School Portal’.

The frame of 'misinterpretation and misuse'

The frame of 'misinterpretation and misuse', present in 400 articles, denounces the detrimental effects of the wrong interpretation and usage of test results, while the importance and potential usefulness of national testing itself remains unquestioned. For example, concern is expressed regarding the risk of interpreting test results as the prime indicator of education quality, while the national tests measure only a set of learning goals.

Particular blame is directed towards politicians and authorities, who are accused of making statements and drawing conclusions that are not statistically sound, thus misinterpreting and misusing the information derived from the tests, sometimes causing 'the warning lights to turn on for no reason' (Lægland, 2016). A second important factor contributing in particular to the misuse of test scores, is the publication of results. This is said to generate competition and enhance a situation in which test results become a goal in itself, rather than a pedagogical tool. Moreover, this competition is said to elicit the adoption of harmful practices aimed at raising test scores at any cost, such as teaching to the test, or abuse of the exemption rule.

The media themselves are blamed for contributing to this competition by awarding test winners and losers, oversimplifying educational debates, stigmatizing schools using 'brutal' and 'distasteful' headings (e.g., Alterhaug, 2016) and caring more about 'newspapers' sales figures' than the impact of their reporting (Traavik, 2007). Moreover, journalists are accused of misleading usage of statistics, which leads to distorted descriptions of the real situation in the educational system. As argued in one article, 'Sometimes a brilliant idea is ruined by poor execution. This may have happened to the national tests in schools - not least by the way that the media have misused this information' (Gårdsvoll, 2005).

As such, this frame conveys the message that the tests themselves are useful tools, while serious problems emerge when test results are misinterpreted, not understood or badly used. The frame thus suggests that the publication of results must be abandoned and that correct interpretation and use of national tests results should be promoted:

If you stop using national tests to rank schools, you will automatically end different practices around exemption. We believe that national tests can have a value in relation to the individual student's learning. Then it is important that students are not taken out for better results. (Ferguson, 2014)

The frame of 'criticism'

The frame of 'criticism', found in 390 articles, conveys a skeptical message towards the usefulness of the national tests, and is highly critical of the detrimental consequences of the test system. Demanded and developed by politicians, rather than school professionals, the national tests are described as adding little to an already 'good selection of mapping tests and other tests that give better insight' (Øgaard, 2017). Even more so, the tests are framed as providing 'a very thin foundation for commenting on learning outcomes' (Skarkerud, 2010). In addition to not living up to the promise of serving as a useful pedagogical and organizational tool, the test system is described as disruptive and as having a negative impact on education quality and equity.

That is, the existence of the national testing system causes a number of negative side-effects. For example, it is argued that the excessive focus on the few measurable educational goals runs the risk of constraining autonomy, diversity and innovation. Similarly, the test system is said to damage students, by reducing motivation, and enhancing conformity of thought, pressure and stigmatization. At the same time, the pressure felt by teachers and principals to obtain high scores is said to lead to the adoption of a number of undesirable and harmful practices. Curriculum narrowing, teaching to the test, excessive test preparation, educational triage and exclusion of low-performing students are particularly denounced:

It has already been revealed that schools drill students in tests similar to the national tests to secure the best possible results. (Ivedestrandsposten, 2005)

A schoolboy was ill on the day of national tests. The principal called home, asked how sick the boy was, brought the boy in his own car to school and then drove him home again after the end of the test. Hi and ho and up the average! (Helland, 2015)

As such, this frame conveys the message that the national tests are not useful, not valid and not reliable. At the same time, the test system hampers good education by causing the adoption of harmful practices and introducing elements into the system with detrimental consequences for education quality and equity. It is therefore recommended that the national testing system is abandoned and that assessment measures are promoted that are considered to be more suitable by school professionals.

Discussion

Informed by framing theory (Entman, 1993), and based on a systematic analysis of 3,046 articles on national tests published in Norwegian regional and local newspapers over a 15-year time span (2004-2018), this paper has identified four dominant discursive frames, which are used by the media when informing readers about national standardized tests. Each frame conveys distinct problem definitions, causal interpretations, moral evaluations and/or treatment recommendations (Entman, 1993).

The analysis revealed differences in the pervasiveness of each frame over time. During the first two test rounds, which were met by severe societal discontent and destructive evaluation reports (Lie et al., 2005), the three frames that express explicit opinions on the test and its use were most pervasive. Upon reintroduction of the tests in 2007, the dominance of these frames declined, which might reflect the more favorable public opinion towards the quality of the tests, and their conditions of administration, thereby giving rise to less debate in the media. At the same time, the frame of 'performance' grew in importance from 2007 onwards and ever since has been the most dominant frame found in Norwegian regional and local press coverage. In contrast to the other three frames, within the frame of 'performance', the existence and validity of the national tests is neither questioned nor debated. Rather, performative-oriented ranking and reporting are common practice, with the potential effect of naturalizing and enhancing competition in education.

While the reasons behind the increasing pervasiveness of the frame of 'performance' are not clear, possible explanations can be advanced that could feed future research. The predominance of this frame may reflect an increasingly widespread acceptance of national tests as valid and useful measures of crucial learning outcomes. The dominance of the frame of performance could also reflect the seductions of quantification (Espeland & Stevens, 2008). That is, indicators such as test scores appear to provide objective, scientific, concrete and politically neutral knowledge of a complex world, thereby appealing 'to the desire for simple, accessible knowledge and to a basic human tendency to see the world in terms of hierarchies of reputation and status' (Merry, 2016, p. 1). This would correspond to a fetish for numbers, measurements and comparisons, as has already been theorized by several educational researchers (e.g., Ozga, 2008; Ball, 2015). Simultaneously, comparisons between entities based on averages, as well as over time, could correspond to the easiest, cheapest or most sensational way for non-technical stakeholders such as journalists, who may not have the expertise to fully understand the complexity of the results (Wu, 2010), to make sense of and report on national tests.

Further research could also look at the impact of different forms of media coverage. To date, most studies have focused on the stress responses of educators in blamed schools (e.g., Elstad, 2009), while less is known about how other media practices, such as praising, rankings and comparisons, which we found to be more common, affect different schools, including those ignored by the media. Such research might be particularly relevant in contexts where schools face increasing competitive pressures and test scores affect the school's reputation. For example, whereas praising may seem like a neutral or harmless media practice, in small municipalities where there are only a few schools, praising and exalting the performance of one or few schools means de facto neglecting and subtly, implicitly blaming the few other local schools. This could therefore correspond to a silent form of 'public shaming and blaming' (Takayama, 2008, p. 388). For neglected schools, the consequences of praising practices might be as harmful as those reported by Elstad (2009) for schools with negative media coverage.

In addition to differences over time, our analysis shed light on interesting geographical differences in the pervasiveness of each of the four frames. Whereas in some counties, such as Oslo and Hordaland, where Norway's biggest cities are located, all four frames are almost equally present, in other counties, such as the more rural Finnmark and Nordland, the frame of 'performance' strongly dominates. While a number of factors could lie behind these geographical differences, such as population size and density, test performance, local media landscapes, local policy contexts, and other cultural specificities, further research is needed to explore them in-depth. While explaining these geographical differences is beyond the scope of this article, they clearly confirm the relevance of not limiting research to national media coverage, which could erroneously lead to the impression of uniformity throughout the country. This might be particularly important considering that previous research has highlighted that media coverage can have an important impact on school actors' experiences and actions (e.g. Elstad, 2009; Stack & Boler, 2007). As such, different media coverage within the same country may have far-reaching effects on school practices and responses to test-based accountability reforms that would otherwise remain difficult to explain. Geographical differences in media coverage therefore deserve a critical spotlight.

References

- Alterhaug, B. (2016, March 17). Nasjonale prøver som gapestokk. [National tests as a pillory]. *Adresseavisen*.

- Arbeidets Rett. (2018, October 14). Tolga skole topper listen over Norges beste skoler. [Tolga School tops the list of Norway's best schools]. *Arbeidets Rett*. <https://www.rett.no/nyheter/skole/tolga/tolga-skole-topper-listen-over-norges-beste-skoler/s/5-44-284096>
- Ball, S. J. (2015). Education, governance and the tyranny of numbers. *Journal of Education Policy*, 30(3), 299–301.
- Baroutsis, A. (2016). Media accounts of school performance: Reinforcing dominant practices of accountability. *Journal of Education Policy*, 31(5), 567-582.
- Baroutsis, A. (2017). Understanding media mentalities and logics: Institutional and journalistic practices, and the reporting of teachers' work. *Discourse Studies in the Cultural Politics of Education*, 40(4), 545-599.
- Baroutsis, A., & Lingard, B. (2017). Counting and comparing school performance: an analysis of media coverage of PISA in Australia, 2000–2014. *Journal of Education Policy*, 32(4), 432-449.
- Borgen, J. H. (2009, December 11). Skolen - der fremtiden skapes. [The school - where the future is created]. *Varden*.
- Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*. doi: 10.1080/02680939.2020.1739337
- Chong, D., & Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science*, 10(1), 103-126.
- Chong, D., & Druckman, J. N. (2010). Dynamic Public Opinion: Communication Effects over Time. *American Political Science Review*, 104(4), 663-680.
- Dixon, R., Arndt, C., Mullers, M., Vakkuri, J., Engblom-Pelkkala, K., & Hood, C. (2013). A Level for Improvement or a Magnet for Blame? Press and Political Responses to International Educational Rankings in Four EU Countries. *Public Administration*, 91(2), 484–505.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173-189.
- Elstad, E. (2012). PISA Debates and Blame Management among the Norwegian Educational Authorities: Press Coverage and Debate Intensity in the Newspapers. *Problems of Education in the 21st Century*, 48, 10–22.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58.

- Eriksen, H. (2011, January 3). Dårligst i landet på 2 + 2. [Worst in the country at 2 + 2]. *Halden Arbeiderblad*. <https://www.ha-halden.no/nyheter/darligst-i-landet-pa-2-2/s/1-2906373-5432755>
- Espeland, W. N., & Stevens, M. L. (2008). A Sociology of Quantification. *European Journal of Sociology*, 49(3), 401-436.
- Ferguson, K. (2014, December 16). Gransker høyt fritak på nasjonale prøver. [Examines high exemption on national tests]. *Bergensavisen*. <https://www.ba.no/nyheter/gransker-hoyt-fritak-pa-nasjonale-prover/s/1-41-7729197>
- Findahl, T. (2017, 30 November). Ingen er bedre til å regne enn denne gjengen. [No one is better at counting than this gang]. *Arbeidets Rett*. <https://www.rekken.no/skole/tolga/nyheter/ingen-er-bedre-til-a-regne-enn-denne-gjengen/s/5-44-206915>
- Frimand, M. (2008, November 13). Stavanger best i klassen. [Stavanger best in class]. *Rogalands Avis*.
- Gamson, W., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 95(1), 1–37.
- Gårdsvoll, S. (2005, December 26). Media misbruker de nasjonale prøvene. [The media abuse the national tests]. *Adresseavisen*.
- Goffman, E. (1974). *Frame analysis: An essay on the organisation of experience*. Cambridge, MA: Harvard University Press.
- Goldstein, R. A. (2010). Imaging the Frame: Media Representations of Teachers, Their Unions, NCLB, and Education Reform. *Educational Policy*, 25(4), 1-34.
- Helland, U. (2015, November 3). Nasjonale prøvelser. [National trials]. *Dagsavisen*. <https://www.dagsavisen.no/debatt/nasjonale-provelser-1.674081>
- Hermstad, G. A. (2008, June 5). Skoletapere i Bindal og Vega. [School losers in Bindal and Vega]. *Brønnøysunds Avis*. <https://www.banett.no/nyheter/article160259.ece>
- House of Lords (2008). The ownership of the news. Volume II: Evidence. London: The Stationary Office Limited.
- Høst, S. (2018). Avisåret 2017 [Newspaper year 2017]. Report no. 86/2018. Volda: Høgskulen i Volda.
- Hovland, P. (2018, November 6). Lena ungdomsskole har funnet formelen. [Lena ungdomsskole has found the formula]. *Oppland Arbeiderblad*.
- Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, 12(1), 19–40.

- Iyengar, S. (1991). *Is anyone responsible?: How television frames political issues*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. London: University of Chicago Press.
- Lie, S., Hopfenbeck, T. N., Ibsen, E., & Turmo, A. (2005). *Nasjonale prøver på ny prøve [National tests retested]*. Oslo: Department of Teacher Education and School Research, University of Oslo.
- Lofotposten. (2013, October 29). Leser dårligere enn før. [Read worse than before]. *Lofotposten*. <https://www.lofotposten.no/lokale-nyheter/leser-darligere-enn-for/s/1-71-6950339>
- Lægland, M. (2016, September 28). Store problemer med presentasjonen av statistikken. [Major problems with the presentation of the statistics]. *Bladet Tromsø*.
- Matthes, J., & Kohring, M. (2008). The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *Journal of Communication*, 58(2), 258–279.
- Macnamara, J. R. (2005). Media Content Analysis: Its Uses, Benefits and Best Practice Methodology. *Asia Pacific Public Relations Journal*, 6(1), 1–34.
- Merry, S. E. (2016). *The seductions of quantification: measuring human rights, gender violence and sex trafficking*. Chicago and London: University of Chicago Press.
- Mockler, N. (2013). Reporting the ‘education revolution’: MySchool.edu.au in the print media. *Discourse: Studies in the Cultural Politics of Education*, 34(1), 1-16.
- Mockler, N. (2020). Ten Years of Print Media Coverage of NAPLAN: A Corpus assisted Assessment. *Australian Review of Applied Linguistics*, 43(3), 117-144.
- Norwegian Media Authority. (2018). Ownership – Media Group. Retrieved from: <https://medietilsynet.no/>
- Øgaard, A. (2017, November 13). Nasjonale prøver, mest til bry. [National tests, mostly to bother]. *Moss Avis*. <https://www.moss-avis.no/skole/debatt/utdanning/nasjonale-prover-mest-til-bry/o/5-67-527611>
- Østbye, H. (2019). Media Landscapes. Norway. Retrieved from: <https://medialandscapes.org/country/norway>
- Ozga, J. (2008). Governing knowledge: Research steering and research quality. *European Educational Research Journal*, 7(3), 261-272.
- Pan, Z. & Kosicki, G. M. (1993). Framing analysis: An approach to news discourse. *Political Communication*, 10(1), 55–75.
- Scheufele, D., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1), 9-20.

- Schmitt, R. (2005). Systematic Metaphor Analysis as a Method of Qualitative Research. *The Qualitative Report*, 10(2), 358-394.
- Shine, K. (2015). Are Australian teachers making the grade? A study of news coverage of NAPLAN testing. *Media International Australia*, 154(1), 25–33.
- Shine, K., & O'Donoghue, T. (2013). Teacher representation in news reporting on standardized testing: A case study from Western Australia. *Educational Studies*, 39(4), 385–398.
- Skarkerud, G. (2010, March 1). Lundstein skole. *Oppland Arbeiderblad*.
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: inconsistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23(1), 5-20.
- Skedsmo, G., & Mausethagen, S. (2016). Accountability policies and educational leadership – a Norwegian perspective. In J. Easley, & P. Tulowitzki, P. (Eds.), *Accountability and Educational Leadership – country perspectives* (pp. 205-223). London: Routledge.
- Skilleås, O. M. (2005, January 24). SV/SU og den hemmelige skolen. [SV / SU and the secret school]. *Bergens Tidende*. <https://www.bt.no/btmeneringer/i/BMKa0/sv-su-og-den-hemmelige-skolen>
- Solvang, O. (2015, October 31). Låner ut lesestoff i skolegården. [Lends reading material in the schoolyard]. *Nordlys*.
- Solvang, O. (2017, November 30). I 2011 lovte de at Nordreisa-skolen skulle bli bedre. Nye tall viser at det er verre enn noen gang. *Nordlys*. <https://www.nordlys.no/skole/nordreisa/storslett/i-2011-lovte-de-at-nordreisa-skolen-skulle-bli-bedre-nye-tall-viser-at-det-er-verre-enn-noen-gang/s/5-34-743731>
- Sørensen, A. E. (2004, February 5). Rot om nasjonale prøver. [Clutter about national tests]. *Rana Blad*.
- Stack, M. (2006). Testing, testing, read all about it: Canadian press coverage of the PISA results. *Canadian Journal of Education*, 29(1), 49-69.
- Stack, M., & Boler, M. (2007). Media and Policy in Education. *Policy Futures in Education*, 5(1), 1-16.
- Statistics Norway. (2019). Population. Retrieved from: <https://www.ssb.no/>
- Takayama, K. (2008). The Politics of International League Tables: PISA in Japan's Achievement Crisis Debate. *Comparative Education*, 44(4), 387–407.
- Traavik, H. (2007, December 28). Skuffende julegave til skolene. [Disappointing Christmas present for schools]. *Bergens Tidende*.

- Tvedestrandsposten. (2005, January 30). Riktig bilde av nivået i skolen? [Correct picture of the level in the school?]. *Tvedestrandsposten*.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, 21(2), 221-237.
- Ulmer, J. B. (2014). Re-framing teacher evaluation discourse in the media: An analysis and narrative-based proposal. *Discourse: Studies in the Cultural Politics of Education*, 37(1), 43–55.
- Waldow, F., Takayama, K. & Sung., Y. K. (2014). Rethinking the Pattern of External Policy Referencing: Media Discourses over the ‘Asian Tigers’ PISA Success in Australia, Germany and South Korea. *Comparative Education*, 50(3), 302-321.
- Wu, M. (2010). Measurement, Sampling, and Equating Errors in Large-Scale Assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.
- Yemini, M., & Gordon, N. (2015). Media Representations of National and International Standardized Testing in the Israeli Education System. *Discourse: Studies in the Cultural Politics of Education*, 40(4), 1–15.

Camphuijsen, M. K. (2020). Coping with performance expectations: towards a deeper understanding of variation in school principals' responses to accountability demands. *Education Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-020-09344-6>

CHAPTER 4

Coping with performance expectations: Towards a deeper understanding of variation in school principals' responses to accountability demands

Introduction

In recent decades, metrics and indicators to evaluate the performance of individuals and organizations have proliferated in different policy domains (Mennicken and Espeland 2019). The spread of neo-liberalism and growing popularity of New Public Management (NPM) have contributed to numbers escaping 'from the relatively restricted toolbox in which they were kept' and moving towards more accountability uses (Piattoeva and Boden 2020, p.4). In the education sector, a growing number of countries has adopted standardized tests to measure the performance of schools and teachers, and to hold educational actors accountable for learning objectives determined at the central level, usually emphasizing core subjects or basic skills (Ball, Junemann, and Santori 2017; Verger, Parcerisa, and Fontdevila 2019). Following the increasing use of performance measures for accountability purposes, scholars have referred to an important shift in the governance of education, which goes from professional to performance-based accountability systems (Lingard, Sellar and Lewis 2017). Key objectives of performance-based accountability (hereafter PBA) are to ensure that educators are responsive to and assume responsibility for achieving centrally-defined learning goals, and to promote data-driven decision-making. Whereas generally aimed at regulating actors' behavior, the specific institutional arrangements and instruments of PBA systems can differ according to various dimensions, including the type and alignment of accountability tools, the nature of accountability consequences, and the conceptions of actors that inform tool selection (Maroy 2015; Maroy and Pons 2019).

Despite the increasing popularity of PBA worldwide, studies report mixed results with regards to the impact of PBA on school organization and pedagogical practices. Whereas some educators seem to adapt their practices to meet accountability expectations, others respond by ignoring, avoiding, resisting or re-writing policy demands, or by relying on quick and visible solutions, rather than on thorough and long-term changes (e.g. Barbana, Dumay and Dupriez 2019; Candido 2020; Diamond and Spillane, 2004; Falabella 2014; Maroy and Pons 2019; Verger, Ferrer-Esteban and Parcerisa 2020). In line with recent scholarship on policy enactment (Ball, Maguire and Braun 2012), these studies

demonstrate how, rather than a linear and top-down process, putting policy into practice forms a creative, complex and sometimes constrained social process. To understand variation in policy responses, it is key to examine how actors perceive, interpret and translate policy demands in various ways, while being attentive to how this sense-making process is guided by educators' pre-existing knowledge, beliefs and practices (Coburn 2001, 2004; Spillane and Jennings 1997), as well as enabled or constrained by contextual factors (Braun, Ball, Maguire and Hoskins 2011).

While research interest in PBA reforms has sparked in recent decades, in particular 'hard' or 'strict' accountability systems have been studied extensively, usually in contexts also characterized by high levels of marketization, such as England, the U.S. and Chile. 'Hard' accountability systems tend to rely on external and closely aligned policy tools as well as high-stakes consequences, following the conception of a utilitarian and strategic actor (Maroy 2015). In contrast, other accountability approaches, including 'soft' and 'reflexive' systems, have remained under-researched (Maroy and Pons 2019). These approaches, informed by the conception of an actor that is socially embedded and reflexive, attempt to target actors' internal feelings of responsibility and reflection. Nonetheless, whereas these approaches aim to instigate change 'from the inside' through influencing actors' dispositions, some systems combine internal measures with external devices and moderate to significant accountability consequences (Maroy 2015). As a result of the predominant focus of accountability research on 'hard' systems, a limited understanding persists of how educators perceive and respond to PBA demands in other contexts, as well as of how potential variation in actors' responses can be explained. With the aim of contributing to this understanding, this paper reports on a study that examines how Norwegian principals reflect on and respond to PBA demands. More specifically, the study addresses the following research questions: (1) In what ways do Norwegian principals perceive, interpret and translate PBA demands?; (2) What are the response patterns employed by principals to address PBA demands?; and (3) What are the social mechanisms and contextual conditions (local accountability regimes, school-specific factors and personal trajectories) that explain the response patterns, and the differences among them?

Norway forms a particularly interesting context for this object of study as the Norwegian approach to PBA differs in important regards from 'hard' accountability systems, while at the same time significant local variation is found in accountability regimes. With recent policy documents placing strong emphasis on learning and basic skills as main missions for schools (Larsen, Møller and Jensen 2020), national tests and value-added models are increasingly used to hold teachers, school leaders, and local

authorities accountable for students' learning outcomes and acquisition of basic skills. Nonetheless, the system remains characterized by a relative absence of material consequences (such as financial rewards or sanctions), as well as low levels of marketization. Rather, the Norwegian approach to PBA combines administrative control devices with institutional regulations aimed at encouraging reflection, self-evaluation and organizational learning, so to ensure educational actors adapt their practices in line with the competency aims formulated in the national curriculum and use achievement data for school improvement purposes (Camphuijsen, Møller and Skedsmo 2020). At the same time, despite these generic features at the central level, significant local variation exists in how accountability plays out in practice, following municipal discretion in terms of accountability tools and consequences, as well as local variation in the role played by external audiences, or 'third-party' account-holders. With regard to the latter, local differences exist in school choice regulations (NSD 2016) and the level of activity of local media outlets in reporting on test performance (Camphuijsen and Levatino submitted). Such local variability makes Norway an excellent case to advance the understanding of how different accountability configurations and local policy contexts mediate policy enactment processes and policy outcomes.

This study focuses specifically on principals, as principals play a key role in the enactment of schools' accountability (Coburn 2004; Diamond and Spillane 2004). While often juggling multiple, and sometimes conflicting, accountability demands from different audiences (Pollock and Winton 2016), principals act as key "managers in the middle" or policy brokers. The ways in which principals reflect on and respond to new policy demands are crucial, not in the least because principals' reflections and actions have the potential of mediating teachers' experiences and responses (Diamond and Spillane 2004; Spillane, Diamond, Burch, Hallett, Jita and Zoltners 2002). Principals furthermore form a particularly interesting group of school-level actors as they, including in Norway, have often been specifically targeted by NPM reforms, expected to act as the vehicles of modernization of education in schools (Møller and Skedsmo 2013).

The analysis of this paper relies primarily on qualitative data derived from 23 in-depth interviews with primary school principals in nine urban municipalities in Norway, characterized by diverging local accountability regimes. In order to examine and explain different response patterns employed in reaction to PBA demands, this study relies on the sociological concept of 'reactivity', understood as the way 'individuals alter their behavior in reaction to being evaluated, observed, or measured' (Espeland and Sauder 2007, p.6). More specifically, the study relies on two social mechanisms

identified by Espeland and Sauder (2007) to understand the reflexive interactions between actors and measures, this to undertake an attempt to explain why principals respond in particular ways to standardized testing and PBA.

The paper is structured as follows. The next section presents a review of previous research on how educators respond to PBA demands. Based on this review, the gap in the existing literature is identified, which this paper tries to address by relying on the ‘reactivity’ framework, outlined in the subsequent section. Thereafter, the Norwegian educational context is briefly explained, which is followed by the study’s methodology. Subsequently, the study’s findings are presented in the form of three predominant responses; (a) alignment; (b) balancing multiple purposes; and (c) symbolic responses. The final section discusses the main results and concludes by arguing that even in the relative absence of material consequences and low levels of marketization, standardized testing and PBA can drive behavioral change, by reframing norms of good educational practice, and by affecting how educators make sense of core aspects of their work.

Literature review: school actors’ responses to PBA

In recent years, a growing body of studies has examined how educators respond to accountability approaches characterized by the ambition to elicit change ‘from the inside’ by influencing actors’ dispositions, as well as a relative absence of material consequences. A key finding of these studies is that policy enactment processes can differ significantly from policy intentions and even contradict key assumptions of the action theory underpinning such PBA reforms. For example, based on a study conducted in three schools in French Belgium, Barbana, Dumay and Dupriez (2019) showed how the clash between the accountability instruments and educators’ own views on instruction and student assessment discouraged many teachers from adopting the anticipated ‘reflexive attitude’ and from making substantive changes to their classroom practices. At the same time, the authors found that a minority of teachers expressed a more positive attitude towards the instruments, and used them to reflect on and to a certain extent modify their practices¹¹ (Barbana et al. 2019). Similar findings were reported in the Brazilian context, where Candido (2020, p.22) found that a number of educators chose to adapt their discourses and practices to testing and accountability policies, while other school actors

¹¹ The authors classified the three schools as ‘atypical’, as they were theoretically selected for their ‘a priori distance towards the worldviews embedded into the accountability instruments’ (Barbana et al. 2019, p.11).

found ways to ‘rewrite the rules of the “game” to fit their own interests’. Also in Norway, studies report mixed results with regards to the impact of standardized testing and PBA demands on educators’ practices. Whereas some studies report how school leaders employ symbolic responses to policy demands emphasizing test scores (Gunnulfson and Møller 2017), other studies show how national testing and PBA have an important impact on instructional strategies and schools’ internal affairs (Elstad 2009; Seland, Vibe and Hovdhaugen 2013; Skedsmo 2018).

A second key contribution of this growing body of literature entails the documentation of side-effects, formerly in particular associated with ‘hard’ accountability approaches (e.g. see Au 2007; Mittleman and Jennings 2018). For example, in both Germany and Israel, where accountability systems were deliberately designed without attaching high-stakes consequences so to avoid the emergence of side effects, scholars report effects such teaching to the test, educational triage, and curriculum narrowing (Feniger, Israeli and Yehuda 2015; Thiel, Schweizer and Bellman 2017). Rather than attributing side-effects to the stakes of accountability, Thiel et al. (2017) suggest that side effects might form systematic problems of accountability in education, while Feniger et al. (2015, p.3) point towards the ‘power of numbers’, arguing that ‘the use of external standardized tests, in itself, causes a shift in the way actors in the educational field think and speak about education’.

So far, a limited understanding prevails of how to interpret and explain the complex, creative and sometimes unanticipated responses adopted in these accountability contexts. That is, little remains known of *why* and *under what circumstances* educators may adopt particular responses. By identifying the social mechanisms that induce particular response patterns, and by establishing the conditions under which they operate, a deeper understanding can be gained of ‘why we observe what we observe’ (Hedström and Svedberg 1998, p.9). With the aim of contributing to this understanding, this study relies on the sociological concept of ‘reactivity’, and more specifically, the framework developed by Espeland and Sauder (2007), which identifies two social mechanisms that induce reactivity.

Reactivity as an analytical device to interpret and explain responses to PBA

In recent decades, awareness has grown that, because people are ‘reflexive beings who continually monitor and interpret the world and adjust their actions accordingly’, social measures such as standardized tests are ‘reactive’ (Espeland and Sauder 2007, p.2). While some see it as a methodological problem that people adapt their actions in response to being measured (Campbell 1957), others

consider reactivity a promise and a vehicle for inducing behavioral changes in desired ways. Considering that PBA systems tend to rely on the latter understanding of reactivity, it is key to examine the reflexive interactions between educators and PBA instruments, in order to gain a deeper understanding of the reactions employed by key actors, as well as the effects they give rise to.

To do so, this study relies on the framework developed by Espeland and Sauder (2007). Based on a large-scale study on law school rankings, they identified two mechanisms that produce reactivity to social measures; self-fulfilling prophecies and commensuration¹². Rather than restricting the definition of a self-fulfilling prophecy to false beliefs (Merton 1968), Espeland and Sauder (2007, p.11) refer to ‘processes by which reactions to social measures confirm the expectations or predictions that are embedded in measures or which increase the validity of the measure by encouraging behavior that conforms to it’. Key in this regard is the understanding of social measures, designed to evaluate the performance of individuals or organizations, as carrying tacit assumptions about what constitutes ‘quality’, ‘excellence’, or ‘success’, thereby reframing or constructing new norms of what is considered relevant, valuable and desirable. By encouraging actors to see themselves and behave according to the norms of good practice embedded in measures, thereby reinforcing their validity, social measures can create self-fulfilling prophecies (Espeland and Sauder 2007). In this light, various scholars have argued how, in the education sector, the increasing use of performance metrics has reoriented the purposes of schooling and redefined the education profession (Ball 2003), while simultaneously showing how educators can come to internalize or embody new definitions of ‘quality’, ‘excellence’ and ‘success’, fostering efforts of norm compliance (Courtney 2014). Another way in which social measures can operate as a self-fulfilling prophecy relates to the effects of measurement on perceptions and actions of external audiences (Espeland and Sauder 2007). In particular when the precise, quantitative distinctions between individuals and institutions are increasingly perceived as relevant and ‘natural’, even statistically insignificant differences can have real consequences for measured objects. That is, when external audiences act upon such differences, for example by raising their voice or choosing another provider, differences that initially largely resulted from measurement noise can become real and strengthen over time.

¹² While Espeland and Sauder (2007) present self-fulfilling prophecies and commensuration as two distinctive mechanisms of reactivity, they simultaneously emphasize how the two mechanisms can interact in the production of behavioral change.

Commensuration, the second social mechanism identified by Espeland and Sauder (2007), entails ‘the comparison of different entities according to a common metric’ (Espeland and Stevens 1998, p.313). Prices constitute an example of commensuration, which have become a highly naturalized form of comparing the value of disparate goods or services. Standardized test scores are another example of commensuration, which also enable the formal comparison of disparate entities, such as schools that are located in different parts of the country, with diverging histories, cultures and student populations. While self-fulfilling prophecies induce behavioral changes as a result of actors adapting their actions in response to altered expectations, commensuration shapes behavior by changing ‘what we pay attention to, which things are connected to other things, and how we express sameness and difference’ (Espeland and Sauder 2007, p.16). One way in which commensuration affects sense-making is by simplifying and de-contextualizing information, while organizing what remains into numbers that often appear rational, objective and robust, and are easy to interpret and quick to compare and disseminate. The more such numbers become taken-for-granted ways to evaluate and compare goods or entities, the more attention risks being diverted from other ways of expressing difference. For example, the more standardized test scores or rankings are perceived and acted upon as proxies of school quality, the more attention shifts away from other differences between institutions, in particular differences that are hard to quantify. Another feature of commensuration that affects sense-making relates to the creation of precise and hierarchical relationships between measured objects, which enables the possibility to compare oneself to others, as well as previous versions of oneself, thereby affecting how entities make sense of one another and themselves, and changing how one determines and assesses ‘progress’. In recent years, it has been suggested that the constitutive power of commensuration can, in part, be attributed to the relation between data and affect (Sellar 2015). As argued by Sellar (2015, p.135), commensuration can shape actors’ experiences and behavior as a result of ‘emotional or felt effects that data and associated judgements have on those whose practices are made commensurate in order to be compared and evaluated, sanctioned and rewarded’. In this regard, scholars have highlighted how performance data, in particular when used to compare and judge individuals or institutions, can engender feelings such as pride, shame and envy (Ball 2003), and as such influence sense-making processes.

Measurement, transparency and accountability in the Norwegian context

In Norway, the increasing presence and regulatory power of external actors concerned with measurement, observation and evaluation has altered historical self-regulatory dynamics of the education profession (Skedsmo and Mausethagen 2016). With the turn of the millennium, growing concerns about below-expected learning outcomes of Norwegian students in basic skills such as reading contributed to the increasing call for external assessment of student performance as well as external control of educators' competence and results (Møller and Skedsmo 2013). In 2004, a national quality assessment system was adopted, which includes national standardized tests. The tests are administrated at the beginning of the 5th, 8th and 9th grade and measure students' acquisition of basic skills in reading and numeracy as well as their performance in English, reflecting competency aims as formulated in the national curriculum at the end of grade 4 and 7. Value-added models have been published at the school-level since 2016, in response to calls for more accurate measurement of schools' contribution to student learning.

The main rationales behind national testing and value-added models are to assess whether schools succeed in teaching pupils centrally-defined learning objectives as well as to foster data-driven decision-making by providing teachers, school leaders and local authorities with student performance data. Following the conception of an actor guided by social obligations, Norway's PBA system relies on a combination of external control devices as well as institutional arrangements that encourage self-evaluation processes and target internal feelings of responsibility and reflection (Camphuijsen, Møller and Skedsmo 2020). More specifically, a main external pressure mechanism forms the publication of test results on the government website 'the School Portal' (skoleporten.no), where school results are presented in the form of comparisons to the municipal, county, and national average, following a benchmarking logic (Skedsmo 2018). In this regard, the media form an important 'third-party' account-holder, by regularly reporting on municipal and school performance, often in the form of performative-oriented rankings and with a focus on narratives of success and failure (Camphuijsen and Levatino submitted; Elstad 2009). Here it is important to emphasize that the extent to which the publication of results plays out as a high-stakes mechanism for schools depends for an important part on the degree of school choice families enjoy.

Administrative supervision conforms to a second external control mechanism. Primary and lower-secondary schools report primarily to the municipal superintendent, who monitors and controls schools' results on various quality assessment measures, including standardized tests. Municipal

authorities simultaneously play an important role in encouraging reflection, self-evaluation and organizational learning on the basis of performance data, by supporting and following-up schools' routines for analyzing and using assessment results to foster school improvement. As highlighted in the introduction, significant local variability exists in municipal routines surrounding administrative supervision, support and follow-up (e.g. Seland et al. 2013; Skedsmo 2018). In addition, municipalities have adopted different practices around the publication of test results on municipal websites and school choice regulations. With regard to the latter, while school choice in compulsory education is generally restricted in Norway, data from the Norwegian Centre for Research Data (NSD) reveal that 56 Norwegian municipalities (i.e. 19.5%) allow for 'freer user choice in the area of compulsory education' (NSD, 2016, p.256)¹³. Still, while such municipal regulations are likely to award families some degree of school choice, priority remains given to students residing in the school's catchment area.

Whereas initially strongly resisted (Tveit 2014), recent studies indicate that over time, standardized tests have become more broadly accepted. In particular school leaders see benefit in having access to performance data (Seland et al. 2013), while teachers continue to struggle with how to integrate test data in their daily practices (Gunnulfson 2017; Mausethagen 2013; Mausethagen, Prøitz and Skedsmo 2017). Nonetheless, most schools have established systems surrounding how to use achievement data for school improvement purposes, in line with policy expectations (Seland et al. 2013).

Data and methodology

This paper relies primarily on qualitative data derived from 23 in-depth interviews with principals of primary schools located in nine Norwegian municipalities. The municipalities are located in eight different counties, dispersed across all regions of the country. Interviews were considered a particularly suitable method to gain a deeper understanding of principals' world-views, motivations and professional trajectories, as well as perceptions, interpretations and translations of PBA demands. Recognizing that principals' perspectives are influenced by local policy contexts and school-specific factors (Braun et al. 2011), I sampled schools with the aim of guaranteeing variability in important

¹³ The data are based on "Survey on Municipal Organization 2016, Municipality File". The survey was financed by the Ministry of Local Government and Modernisation (KMD). The data are provided by Norwegian Institute for Urban and Regional Research (NIBR), and prepared and made available by NSD – Norwegian Centre for Research Data. Neither NIBR, KMD nor NSD are responsible for the analyses/interpretation of the data presented here.

factors, while ensuring comparability in others. With regards to the latter, I decided to focus specifically on public schools, which enroll 96% of the country's student population at the level of compulsory education (Statistics Norway 2019). Additionally, all selected schools are located in urban municipalities. Compared to their rural counterparts, urban municipalities in Norway tend to possess greater institutional capacity to assist schools in policy enactment. As this study had a specific interest in the role of local authorities in supervising and supporting quality improvement efforts, institutional capacity was considered an important requirement.

For the selection of schools, I first classified Norway's urban municipalities according to two criteria, with the aim of assuring variability in local policy contexts. The first criterion entailed the level of 'strictness' of the local accountability regime (i.e. whether municipalities had a strong or weak performance orientation, as well as the type and level of alignment of the local accountability instruments). To determine local accountability regimes, previous research, local policy documents and survey data were used. The latter data were collected during the school year of 2018-2019, in the context of the 'REFORMED' research project. The second criterion referred to the level of involvement of 'third-party' account-holders (i.e. parents and the media). This criterion was made up of two combined variables: first, whether municipalities employed freer or restricted regulations around school choice, and second, the level of activity of the local media in reporting on national testing. While I used NSD data (2016) to secure local variability in regulations around school choice, I relied on a unique database on local media coverage of national testing (2004-2018) to determine the level of activity of the local press. This database reveals important local differences in the extent to which schools are named, blamed and praised for performance in local and regional newspapers (Camphuijsen and Levatino submitted).

Following this classification of municipalities, primary schools were selected in different local policy contexts. For the selection of schools, secondary data provided by the Norwegian Directorate for Education and Training, as well as survey data derived from the administration of the electronic questionnaire in the context of the 'REFORMED' research project were used. More specifically, the sampling criteria for the selection of schools included: a) school performance; b) reported performative pressure levels; c) reported reputation and d) reported pressure to maintain enrolment perceived by principals (see Ferrer-Esteban 2020). Ideally, a proxy of the schools' social composition would have been used as a fifth sampling criterion, as previous research has shown that student composition can influence how PBA policies, as well as school choice regulations, are experienced

and enacted at the school level, in part because composition tends to correlate with school performance (e.g. Keddie 2013). Nonetheless, in the absence of available secondary data on Norwegian schools' social composition, this was not possible. However, the interviews with principals, who described the dominant socio-economic status (SES) of their school's student population, as well as the percentage of minority-language students, confirmed that the sample is characterized by significant variability in the schools' social composition, as shown in Table 1. The sample of principals, moreover, is characterized by variety in gender (12 female principals and 11 male principals) and years of experience working as a school leader (ranging from four to 26 years). All principals worked as teachers before taking on the position of school leader, as is common in Norway, while 19 of them had obtained or were in the process of obtaining formal education in school leadership.

Table 1. Overview of participating schools

School ID	Student composition		Performance
	Predominant SES status	Percentage of minority language students	
School 1	Low	> 75%	Low
School 2	Low	> 75%	Low
School 3	Low	50-75 %	Low
School 4	Low	25-50 %	High
School 5	Low	25-50 %	Low
School 6	Average-low	> 75 %	High
School 7	Average-low	25-50 %	Medium
School 8	Average	50-75%	Low
School 9	Mixed	50-75%	Medium
School 10	Mixed	25-50%	Medium
School 11	Mixed	< 25%	Medium
School 12	Mixed	< 25%	Low
School 13	Mixed	< 25%	Medium
School 14	Mixed	< 25%	Medium
School 15	Average-high	< 25%	Medium
School 16	Average-high	< 25%	Low
School 17	Average-high	< 25%	Medium
School 18	Average-high	< 25%	High
School 19	High	< 25%	High
School 20	High	< 25%	High
School 21	High	< 25%	Medium
School 22	High	< 25%	High
School 23	High	< 25%	High

Data collection was informed by the policy enactment perspective (Ball et al. 2012; Braun et al. 2011), which recognizes that, rather than a straightforward, linear and mechanical process, responding to policy demands forms a dynamic, non-linear and negotiated process. Key actors at different levels

(e.g. municipality, school, and classroom) are involved in interpreting and translating abstract policy ideas in complex and creative ways, enabled and constrained by local contexts and school-specific factors. As such, the interviews focused on gaining an understanding of how principals perceived, interpreted and translated policy demands emphasizing data-driven decision-making and PBA, as well as the mediating role of personal and contextual factors. Specifically, the individual interviews, which were conducted between May 2019 and March 2020, followed a semi-structured interview script, which addressed (a) biographical information; (b) school characteristics and school context; (c) interpretations of testing, transparency and accountability demands; (d) performative pressure; (e) pedagogical practices and data-use (f) administrative accountability. All interviews, which lasted on average 50 minutes, were taped and transcribed verbatim.

The data analysis consisted of three phases. During the first phase, I performed a reading of all 'raw' interview transcripts, in order to obtain a holistic view of the themes emerging during the interviews. During the second phase, a codebook was developed, which emerged in part from this first reading, as well as built on the heuristic distinction between policy interpretation and translation developed by Ball et al. (2012). The coding of all interviews, for which qualitative data analysis software was used (Atlas.ti), and the subsequent data analysis, allowed for the identification of three distinct response patterns in how principals perceive, interpret and translate policy demands emphasizing data-use and PBA. The following overarching codes were particularly significant in this regard: (a) importance awarded to test results; (b) perception of PBA and data-use; (c) experience of performative pressure; (d) strategies to secure achievement of basic skills. The final phase aimed at making sense of and explaining the three response patterns, for which I relied on the reactivity framework. During this final reading, each response pattern was examined in further depth, and it was analyzed how different manifestations of the social mechanisms identified by Espeland and Sauder (2007) served as explanatory factors in interpreting the responses. It was moreover examined how the different manifestations of the two mechanisms interacted with contextual and personal factors in inducing each particular response pattern.

Findings

The presentation of the study's findings is structured according to the three main response patterns articulated by principals. However, before describing the different response patterns, it is important

to highlight that the analysis identified a number of similarities in principals' interpretations and translations of policy demands emphasizing data-driven decision-making and PBA. It appeared that all principals, including those most critical of standardized testing and PBA, consider it an important school mission to make sure that students achieve the basic skills established in the national curriculum. Similarly, all principals report to have established routines for preparing students for the tests¹⁴, as well as routines for analyzing test results. The most common use of test scores reported by principals is to identify students in need of support and follow-up, which all mention to do.

Despite these similarities, principals' perceptions and interpretations differ, both with regards to their general conceptions of the path to school improvement, and their perceptions and interpretations of standardized testing and PBA. With regards to the latter, it was found that principals' perceptions diverge regarding whether they perceive national tests as valid, fair and useful measurements, as well as how much importance they ascribe to the achievement of basic skills compared to other educational goals and purposes (i.e. how much emphasis they place on academic achievement). The analysis reveals that these different perceptions and interpretations guide principals' translations, generating three distinct response patterns, which I named (a) alignment; (b) balancing multiple purposes; and (c) symbolic responses.

The first pattern is articulated by principals who perceive the national tests as valid measures of crucial competencies, and aim at obtaining the best possible test results by adopting top-down, performance-oriented management practices and data-driven decision-making. In other words, they align their practices to accountability expectations. The second pattern is employed by principals who also perceive the national tests as valid measures, but who, in contrast to principals in the first response category, reject a prioritization of the tested cognitive competencies over non-cognitive competencies. Moreover, rather than the main source of information, the tests form one out of various information sources that guide school development. The third pattern is articulated by principals who question the validity of the national tests, and strongly emphasize non-cognitive competencies as key educational purposes. They respond predominantly symbolically to the expectation to raise test scores and employ data-driven decision-making. The three response patterns are summarized in Table 2, and explained in further depth below. For each pattern, it is highlighted how principals interpret and translate policy

¹⁴ The main rationale behind preparing students for the test is to make sure students are familiar with the test format and test situation. Such preparations are recommended by the Norwegian Directorate for Education and Training, and are perceived particularly important for primary schools, where students have little to no experience with taking tests.

demands, following the heuristic distinction between interpretation and translation developed by Ball et al. (2012). Interpretation refers to the initial reading of policy texts, a process during which actors construct meaning of policy ideas and attempt to make sense of policy demands. Translation relates to the language of practice, and the decisions made by actors regarding how to put into practice abstract policy ideas (Ball et al., 2012). In addition, it is discussed how different manifestations of the social mechanisms outlined in the theoretical framework interact with contextual and personal factors in inducing the response patterns.

Table 2. Overview of three response patterns

Response pattern		Interpretation	Translation	Mechanism of change
	Perception of national tests and PBA	General conception of path to school improvement		
1. Alignment Strong commitment to PBA mandate and action theory	Tests are valid measures of education quality Tests measure crucial competencies Tests provide basis for school management and improvement of school practice Transparency and accountability for test results are positive and natural	Belief in school improvement by means of top-down, performance-oriented management and data-driven decision-making	Ambition to obtain highest possible academic performance defines the school vision and organizational goals Data analysis is highly routinized, and serves to diagnose the schools' situation and change practices according to "what works"	Self-fulfilling prophecy: expectations embedded in standardized testing and PBA system are internalized and attempted to be realized
2. Balancing multiple purposes Partial/selective commitment to PBA mandate and action theory	Tests are valid measures of education quality Tests measure important competencies, but non-tested competencies are equally important Tests represent one out of various information sources	Belief in school improvement by means of collective dialogue and reflection, using test results as one out of various sources of information	A balanced focus on academic and social development defines the school vision and organizational goals Data analysis is highly routinized, and serves as a source for collective reflection and decision-making	Commensuration: standardized testing creates a precise hierarchy of schools, which shapes schools' self-perceptions and guides decision-making

	that inform school development			
	Concerns about harmful effects of transparency			
3. Symbolic responses	Concerns about tests' validity	Belief in school improvement by means of continuous improvement of classroom practices based on educators' professional knowledge and judgement	A balanced focus on academic and social development defines the school vision and organizational goals	Commensuration: test scores are increasingly perceived as prime indicators to evaluate, judge and compare school quality
Weak commitment to PBA mandate and action theory	Tests measure important competencies, but non-tested competencies are equally important		Data-analysis forms a less routinized and more isolated practice, with a limited impact on school development work	Self-fulfilling prophecy: this perception can generate pressure from external audiences to conform to the definitions of good practice embedded in standardized testing and PBA
	Tests serve as external control, and provide information of limited use for school development			
	Concerns about harmful effects of transparency and accountability			

Alignment: Achievement of basic skills as core school mission

Interpretation

The first type of response is articulated by principals who strongly align to the PBA mandate and action theory. Principals in this response category perceive basic skills as crucial competencies that students need in order to obtain further education and improve their life chances. One principal captured the dominant view by arguing that ‘if children cannot read, write and calculate, then they do not manage their life’ (P1)¹⁵. Inspired by rhetoric often promoted by research on school effectiveness and visible learning, principals express a strong belief in the ability of educators to impact student learning, regardless of a student’s family background. Three of the four principals who more explicitly adopt this approach work at schools that perform (significantly) below municipal and national averages (P1; P2; P12). In two of these cases, the schools’ student populations are characterized by a high proportion of students with a minority background and low SES status. Despite working with

¹⁵ To assure the anonymity of the research participants, each interviewee is referred to by a numerical code. The letter ‘P’ stands for principal, while the number in the coding refers to the school ID (see Table 1). All interview quotes have been translated from Norwegian to English by the author.

more disadvantaged student populations, principals are strongly motivated by a ‘no-excuses motto’ and work hard to raise expectations of students among teachers. The fourth principal in this response category works at a school that performs at the municipal and national average, with a student population characterized by a higher number of students with more advantaged backgrounds (P15).

All four principals describe the fact of having to take responsibility and being held accountable for students’ learning outcomes as ‘natural’ (P12) and ‘positive’ (P1). They generally support the publication of results, referring to transparency as an important trigger of school improvement¹⁶. In a similar vein, the principals tend to speak positively about administrative control exercised by the municipal superintendent to secure obtainment of basic skills:

So, I am happy that [municipality X] is a very demanding municipality. It does not hold back. They are forward-leaning in almost everything they do, [they] make demands on their schools. I was told that I got three years to get good results at this school. My boss said it clearly. And that is how I like it. (P1).

Even though principals highlight the importance of administrative pressure to ensure school actors take responsibility for academic results, they simultaneously explain that the pressure they themselves experience is either for an important part or fully self-imposed:

It is completely clear to them [local authorities] that we have a worse starting point and that things take more time. They have an understanding for that. We are not being compared to [neighborhood X, a very affluent neighborhood], even though we like to compare ourselves to [neighborhood X]. Our students will be just as successful as those in [neighborhood X]. This is actually our motto, to “out-compete” those in [neighborhood X]. That is the goal. [...]. There is nobody who gets angry because there are bad results, but we often become very disappointed ourselves. (P2).

We probably do it to ourselves, I think. [...]. I do not think there are many principals who have no ambition to do well, so you put some pressure on yourself to do well and try to find the methods to boost results. After all, I have to admit, if I had bad results, I would feel that I have not delivered, I would go home with stomach ache. It touches me. (P15).

The pressure that principals impose on themselves, as exemplified by the quotes above, seems to derive for an important part from an internal sense of duty to bring about high academic achievement, and not to let their students down by allowing for sub-optimal performance. As test performance is more and more perceived as a proxy of the extent to which principals succeed in giving their students a chance to get ahead in life, principals increasingly measure their professional success against external metrics, and value themselves according to the progress their school makes in moving up in the

¹⁶ Principals acknowledge that, if misused, the publication of results can have negative side-effects. Nonetheless, not publishing test results is still perceived as more problematic.

performance hierarchy. This exemplifies how principals can come to internalize norms embedded in standardized testing and the PBA system as to what counts as relevant purposes of education. In the words of one principal:

[The tests] give us a good clue as to what teaching should contain. Not that we should teach to the test, but it does give us some guidelines regarding what to emphasize, what is important for kids to learn. (P12).

Translation

Principals' actions, in turn, are steered by the belief that aiming at the best possible test results forms a key way to secure educational "success". In an attempt to do so, principals adopt top-down, performance-oriented management practices and data-driven decision-making. More specifically, academic achievement is formulated as a core organizational goal, sometimes in the form of specific performance targets (e.g. the ambition to perform at or above the municipal average), which guide the schools' organizational, instructional and pedagogical approaches. In particular in the case of the two lowest-performing schools (P1; P2), the perceived importance of a strong focus on academic achievement seems to contribute to the degradation of other organizational goals and identities. For example, after being newly appointed at a school that for years had performed significantly below the national average, the principal of school 1 decided to build-down the former school vision, which strongly emphasized relevance for students' multicultural identity. In contrast, the new vision that was created more strongly emphasizes academic achievement, a decision with significant consequences for how the school is run:

We have laid far below the national average. So, my ambition was to turn the school around. [To] think less about the multicultural and more about the academic content. We have been working a lot on classroom structure ... I have replaced half the school staff. [I have replaced] the entire management. So here have been major upheavals. (P1)

The four principals explain that at their schools, data-analysis efforts are highly routinized and test data are used for multiple purposes, including when distributing resources, assessing and comparing teachers' performance, moving around staff members, and deciding on the school's focus areas and improvement plans. The four principals, moreover, show strong commitment to the use of achievement data to identify learning gaps and find out "what works", i.e. what methods and practices result in higher student learning. As explained by one principal, this entails constant reflections on questions such as 'could we have done this differently, could we have done better, could we have

worked differently so that more students would have understood the tasks?’ (P2). In addition to identifying methods and practices that “work”, data-analysis efforts focus on assessing what particular content areas should receive more attention, so to adjust teaching accordingly.

Mechanism of change

As exemplified above, the four principals seem to have accepted and internalized the conception of education as embedded in standardized testing and the PBA system, which stipulates that the key towards educational “success” or “excellence” lies in securing optimal performance in tested skills. In their efforts, the four principals largely conform to these assumptions, and increasingly attempt to become like the definition of “a successful school” embedded in standardized testing and PBA, sometimes at the expense of other educational purposes and identities. The above forms an example of how standardized testing and PBA, by imposing a particular definition of education and encouraging actors to see themselves and behave according to the norms of good practice, can create self-fulfilling prophecies (cf. Espeland and Sauder 2007).

Various factors are likely to promote principals’ internalization of external performance criteria, including the local accountability regime. The four principals work in three municipalities, all of which are characterized by a strong performance orientation. In all three cases, principals report that the municipal superintendent aims at belonging to the top-performing municipalities in Norway, even though two municipalities have a relatively high number of disadvantaged students. To secure this, the three municipalities have a long history of PBA demands, and rely on an extensive toolbox that both controls schools and serves to promote organizational learning and reflection. The extensive use of PBA tools, as well as their strong alignment (cf. Maroy 2015), can contribute to a process of ‘socialization’. That is, a climate where implicit and explicit expectations emphasize a focus on improving learning outcomes and push schools to conform to expectations surrounding datafication, can come to shape educators’ identities and discipline school practices.

Still, whether principals working in such local contexts actually come to perceive test scores as proxies of professional success seems to depend on their views on education, which are shaped by their professional trajectories. In this regard, the analysis hints at the importance of the educational institution where principals obtained formal education in school leadership. It appeared that principals who had obtained their school leadership degree at the Norwegian Business School more often tended

to evaluate their professional “success” according to performance metrics, compared to principals who studied at a pedagogical institute of a public university. From the interview data, it remains however unclear whether principals already held particular views on education before they applied to the Norwegian Business School, or whether they developed these views and perceptions during their leadership training or even later on.

Balancing multiple purposes: Performance data to inform decision-making

Interpretation

The second response pattern is employed by principals who only partially embrace the PBA mandate and action theory. Principals in this response category ascribe significant importance to the achievement of basic skills, while simultaneously placing important emphasis on non-cognitive skills and social competencies as key educational purposes. A prominent sentiment among these principals is that the Norwegian school has never been only about learning, but equally about ‘Bildung’. The principals who more explicitly adopt this approach work at schools that vary significantly from one another, both in terms of performance and social composition¹⁷. What binds them together, more than an overlap in school-specific characteristics, is principals’ belief in the validity of the national tests, as well as their positive attitude towards the use of achievement data for school improvement purposes. In this regard, principals argue that the quality of the national tests has improved over time¹⁸, and explain that the tests respond to an important need for measures to assess whether schools ‘succeed with teaching’ (P20).

National tests, I think, are a tool. After all, we need some assessments that can help us along the way. I think national tests help us to see how we are doing. And the tests have become much better. [...]. We have stopping points once a year, where we look at the academic results, but not in relation to whether the individual student has done well, but more in terms of how we do as a school. [...]. So, we use the results to reflect; "is this where we want to be"? “Do we feel satisfied with this?” (P10).

The tests are so good, of such high quality, that I have no problem to account for them. The texts are good, the questions are good, and they really reveal what is important. Also, it is very nice to have a benchmark "Do we reach what we aim for?" "Do we manage to have a positive development?" [...]. It gives us an indication of whether we manage to work systematically enough. (P8).

¹⁷ This response pattern was found to be the most commonly applied pattern. Principals who more explicitly adopt this approach include P4; P6; P7; P8; P9; P10; P13; P17; P18; P19; P20; P21; and P23.

¹⁸ Principals are in particular positive about recent changes to the test format, which allow for comparisons over time.

In other words, the tests are said to offer a ‘meta-view’ (P8), which provides feedback to schools and principals about ‘where we are’ (P20). Principals regard this feedback as an important source of information to foster collective reflection and school improvement.

The principals furthermore report that there are particular ‘expectations outside of the house on how to perform on the national tests’ (P10), expressed by local authorities and politicians, and by some parents¹⁹. Generally, principals explain that, even though they are regularly followed-up for results and asked to explain and justify performance in particular by the municipal superintendent, such encounters are not experienced as threatening. Rather, such meetings are often described as arenas to gain ideas, get advice, and request support. Nonetheless, the construction of performance rankings, as sometimes presented during collective meetings with the municipal superintendent or in the media, seems to spark some emotional response by a number of principals. That is, such rankings are mentioned to elicit feelings of pride and shame, contributing to pressure and a desire to score well. Moreover, in particular negative publicity following low performance is said to affect some teachers’ and students’ confidence and motivation. The publication of results is therefore often referred to as doing more harm than good.

Translation

When looking more closely at the language of practice, it appears that principals’ belief in the validity of the tests as well as their desire to use test data as a pedagogical and organizational tool results in highly routinized data-analysis efforts. Interestingly, even though principals express a critical attitude towards the construction of performance rankings, in particular by media outlets, the analysis highlights how for many principals, the performance of their school compared to others, as well as to their own previous performance, forms an important reference point to reflect on and reassure their school is ‘on track’:

¹⁹ The interview data reveal large differences in the extent to which parents are perceived as putting pressure on schools and educators to obtain high test scores. Many principals argue that parents are generally more concerned with whether their child has a good time at school, than with academic results. If they do express an interest in the latter, this tends to reflect a desire to find out how their own child performs, not necessarily the school as a whole. Nonetheless, at some schools, principals argue that parents show increasing interest in school performance, and raise their voice in case of below-expected performance. In extreme cases, test scores are used when “shopping” for schools (P7), even though opportunities to do so remain limited. That is, even in municipalities that employ freer regulations around school choice, priority remains given to students residing in the school’s catchment area.

The most important thing about the national tests is internal use, it is an internal medicine. And in that context, it is ok to compare yourself to other schools and other municipalities to see "are we in the right place, are our students weaker than others?". This can provide knowledge about the need to put in extra support in some areas. (P17).

Based on collective reflections on comparisons of schools' relative performances, principals determine whether there exists a need to adapt organizational, pedagogical or instructional practices. For example, reflection on test performance can contribute to the decision to revisit the distribution of resources, or to offer teachers to take part in professional development courses. Moreover, if it appears that tested students do not manage specific tasks, particular content can be given more emphasis, intensive courses for low-performing students are sometimes set up, or adaptations to teaching methods are considered. The principals explain that some of the implemented changes only account for the grade-level of the tested students, while other changes are more systematic, and imply changes at non-tested grade-levels. From this it appears that, depending on the perceived need to implement changes, the focus on test performance can have an important impact on the school's core activities and practices. Nonetheless, change initiatives remain based on dialogue and collective agreement between school leaders and teachers, and test data form one out of various sources of information that guide principals' decision-making processes. Moreover, in contrast to principals in the first response category, the principals in this response category reject a prioritization of academic performance over other competencies, and continue to place significant emphasis on non-cognitive skills and social competencies in their development projects and focus areas, even in the case of low academic performance.

Mechanism of change

In contrast to principals in the first response category, principals in the second category only partially agree with the conception of "a successful school" embedded in standardized testing and PBA, and continue to balance multiple educational purposes. Instead, the new social relations constructed by commensuration form a more central explanatory factor to understand principals' responses. That is, standardized testing, as an example of commensuration, creates a precise hierarchy of schools, which seems to affect how principals evaluate how their school is faring, as they increasingly rely on performance comparisons to assess 'progress' and assure their school is 'on track'. As such, the perceived need to undertake action is increasingly shaped by a school's relative performance, where

previous performance, or the performance of other schools or the municipality as a whole, serve as reference points.

At first glance, it appears that performance comparisons contribute in particular at low-performing schools to a perceived need to adapt practices so to move up in the performance hierarchy. It is also the principals of these schools that express particular discomfort with the construction of performance rankings by some municipal superintendents or media outlets, which further contribute to a desire to score well, so to avoid public humiliation. Nonetheless, upon deeper examination, it appears that also principals of average-performing schools or even those at the top of the performance hierarchy can perceive a need to adapt their organizational, instructional or pedagogical practices as a result of performance comparisons. That is, many principals seem to evaluate their performance in relation to performance expectations, taking their student population into account. As such, also average- and high-performing schools can perceive a need to undertake action, for example when their relative performance drops or when they perform below what can be expected from their student body.

Symbolic responses: Value-driven education and a focus on the non-measurable

Interpretation

The third response pattern is employed by principals who show weak commitment to the PBA mandate and action theory. These principals tend to more actively question the validity of the national tests²⁰, while simultaneously argue that the tests form a narrow measure of what the schools' priorities are. Principals' concerns about the tests' validity reflect more general skepticism regarding the often-proclaimed superiority of standardized, quantitative data over particularistic ways of knowing. Instead, principals emphasize the importance of professional judgement and knowledge in fostering school improvement. Moreover, rather than a predominant focus on performance and academic excellence, principals support a broader approach to learning, and emphasize the humanistic aims of education, arguing that 'if we get bad at that, it is dramatic, both democratic and socially' (P4). Three of the six principals who more explicitly adopt this approach work at schools that perform below municipal and

²⁰ Principals express concerns about the validity of the national tests on multiple grounds. For example, one principal explains that it is not realistic to demand from 5th-graders to sit still for 90 minutes to conduct a test. As students struggle to do so, the tests are perceived to measure 'who has best prepared their students to sit as quietly as possible' (P5), rather than what students have learned or can do. Other principals argue the tests predominantly reflect students' backgrounds, rather than the quality of the school's teaching practices.

national averages (P3, P5, P16), while the other three principals work at schools that perform at or above average (P11, P14, P22). A common sentiment expressed by all six principals is that, rather than imposing pressure on oneself to obtain high academic results, principals feel pressured to ‘make sure all students feel safe and have a good time at school’ (P3).

As the principals in this response category generally reject the norms embedded in the tests and PBA system as to what constitutes quality education, they express particular concern about how public and political debates on school quality have narrowed to a discussion of test scores:

[The tests] show a small part of the picture. I feel that maybe it has become a little too much, that everything is measured only by them. It seems like the only thing people talk about, in a way, how the school performs at national tests. [...]. Clearly it is important that pupils achieve basic skills, but there are many other things that are important here as well. [...]. I think that the curriculum has swung and that it gets far too much attention compared to everything else that we have to do as well. (P16).

The six principals report that in particular local politicians, the media and some parents increasingly rely on test scores to evaluate and compare schools, sometimes drawing quick conclusions based on what principals refer to as uncritical readings of narrow, quantitative data. As a result, principals worry that even small and insignificant differences in performance can have real consequences for schools, when external actors will use this as a basis to question and interfere with the school’s educational project and pedagogical approach:

We, according to the value-added model, are such a school: we perform well, but according to the student base we have, we should be able to perform better, if you read those numbers a bit uncritically. And then I got a bit worried about it because we see that they take them... I got kind of nervous because politicians read numbers very easily and just look at some tables, and think "OK, but then we do this and that". But when Statistics Norway explains how to read the value-added model, they provide a report of over 100 pages. (P22).

When we have good results, they are out to tell us that we should have had better results, with this local neighborhood. Then I think, "No, we cannot start with those discussions". But of course, I am concerned that we have to deliver good enough. We have to deliver such good results on national tests that we do not get pressured, right? Because the moment we start delivering really poorly, then we will, then we will start to get critical ..., then our entire pedagogy will be put under pressure. (P11).

Translation

Even though the fear of being called out or told what to do by external actors motivates principals to act, they strongly disapprove of what they refer to as short-sighted solutions to quickly raise test scores,

such as spending more time on particular test content when students score low on specific tasks. As argued by one principal, ‘it might be that it is not more mathematics that you should have when you are bad at math. It might be that you need more arts and crafts, or more physical education’ (P11). Principals acknowledge that schools adopting such ‘quick fixes’ are likely to see an increase in their test scores in the short-term, but at the same time emphasize that in the long-run ‘they fall like a rock, because they have not worked on what underpins the results’ (P11). Rather than prioritizing academic performance over other educational purposes, the vision statements of the six schools highlight a similar focus on inclusion, solidarity and creativity as core values, while practical esthetical subjects occupy a central place in the schools’ development projects. Moreover, in contrast to principals in the first response category, the notion of setting specific performance targets is dismissed, and principals are critical of efforts to look for straightforward solutions to complex learning problems. Rather, they argue, focus should lie on continuous and steady improvement of classroom practices.

While the use of national test data is not completely rejected, the tests are generally regarded as offering too narrow, too limited, and too unreliable information to be used when making important school decisions. Rather, principals report that the national tests are predominantly used to identify students in need of additional support and follow-up. In contrast to the other schools, data-analysis efforts seem less routinized and appear to form a more isolated practice. In some cases, principals themselves do not take part in data analysis meetings, but leave these efforts to a team of teachers. Whereas some of the principals appear to perceive some value in using test data to identify learning gaps, for others it seems primarily a way to comply with institutional expectations surrounding data-use. While they, in line with policy expectations, have established systems for analyzing test scores, they continue to express much greater faith in and report to rely on other assessment measures as well as teachers’ judgement and knowledge to foster school development.

Despite that such predominant symbolic responses allow principals to keep running their schools as they see most appropriate, in particular at those schools where performance swings from year to year, or at schools that perform below what can be expected from their student population, the perceived threat of increased pressure from local politicians, as well as complains by parents, forms an important source of concern, which sometimes forces principals to respond in ways that challenge their own principle beliefs and values:

When the Knowledge Promotion²¹ and all this came, with more focus on results and stronger steering, we noticed that with the new grade-levels, we began to retreat. We did not dare to be as progressive [as before]. We did not get any better results, on the contrary, but we were tricked into it. [...]. We did the same things, but we did a little less, because we had to do a little more of this. So then you end up training kids on what they are not good at, instead of cultivating what they are good at, and then you try to do the rest afterwards. (P11).

Mechanism of change

In contrast to the principals in the first response category, who imposed pressure on themselves to obtain high test scores, the six principals in this response category predominantly experience external or socially-imposed pressure to aim at high test performance. This pressure is exercised by local politicians, the media and some parents, following the increasing perception that test scores form prime indicators to evaluate, judge and compare school quality. This perception is influenced by features of commensuration. That is, by simplifying and condensing information, while making it seem more authoritative, forms of commensuration (i.e. standardized tests) attract attention (Espeland and Sauder 2007). The more attention is paid to test performance, the more differences between entities become predominantly expressed by the interval on the shared performance metric. Other ways of distinguishing between entities become, in turn, less salient. For example, as explained by the principals, media reports on test performance frequently use headings such as ‘the best schools’, while contextual information is often erased from such reports, and little attention is paid to whether data are fit for cross-comparisons. As such, the presentation of test scores appears robust and definite, and is increasingly used for making general claims about the overall quality of schools.

As outlined above, principals are in most cases able to keep running their schools as they see fit, regardless of pressure exercised by external audiences. With the exception of one, all principals work in municipalities characterized by a weak performance orientation, and loosely coupled accountability tools. This is likely to provide them greater leeway to respond predominantly symbolically to the expectation to raise test scores and use achievement data to foster school improvement. Nonetheless, regardless of the weaker performance orientation of municipal superintendents, pressure exercised by local politicians, the media and parents can sometimes force principals to respond in ways that go against their own views on education. These cases highlight how the increasing perception of local

²¹ Name of the curriculum reform introduced in 2006, which transformed how Norwegian schools were governed.

politicians, the media, and some parents that test scores from relevant and robust indicators of school quality, shaped by features of commensuration, can reinforce self-fulfilling prophecies, by generating social pressure to conform to the definitions of good practice embedded in standardized testing and PBA.

Discussion and conclusion

This study has examined how primary school principals reflect on and respond to being measured, compared and held accountable for school performance on national tests. The findings highlight three distinct response patterns in how principals perceive, interpret and translate policy demands emphasizing data-driven decision-making and PBA. Principals who strongly align to the PBA mandate and ‘theory of action’ attempt to secure optimal performance by adopting performance-oriented management practices and by using achievement data to plan school improvement. In contrast, principals who only partially embrace the PBA mandate and action theory seem to respond by balancing multiple purposes. That is, principals report to rely on performance data to identify learning gaps and reflect on pedagogical and organizational challenges, but reject a prioritization of academic performance over non-cognitive competencies. Finally, principals who experience a significant mismatch between their views on education and the central policy demands predominantly employ symbolic responses. While they comply with the institutional expectation to use test data, this appears to remain a more isolated practice from their core activities.

By examining the three response patterns through the lens of two social mechanisms identified by Espeland and Sauder (2007), this study has attempted to explain why principals respond in particular ways to standardized testing and PBA. The findings indicate that different manifestations of self-fulfilling prophecies and commensuration form important explanatory factors to understand how standardized testing and PBA can give rise to complex, creative and sometimes unanticipated responses. The study simultaneously highlights how the mechanisms are more likely to be activated under particular conditions, which relate both to principals’ trajectories and views on education, as well as to school-specific characteristics, such as the school’s relative performance level, and parental expectations. Finally, the level of ‘strictness’ of the local accountability regime, as well as the level of activity of the local press, are found to play an important role.

More specifically, the study's findings show how standardized testing and PBA can operate as a self-fulfilling prophecy when principals come to see themselves and act according to the criteria of good practice embedded in the tests and PBA system. This implies that measures, such as standardized tests, increasingly create what they are meant to describe. It was found that the local accountability regime, and in particular the history of PBA demands and the degree of alignment between PBA tools, can favor principals' internalization of performance metrics as proxies of professional success. Nonetheless, whether principals actually do so, seems to depend on their views on education, which are influenced by their professional trajectories. In contrast to previous research, which shows that the point of entry to the teaching profession affects principals' desire to be seen as successful according to system metrics (see Heffernan 2017), this study's findings hint at that for Norwegian principals the institution where they obtained formal education in school leadership, that being the Norwegian Business School or a pedagogical institute at a public university, matters more than the entry point to the profession. Nonetheless, from the interview data, it is not possible to draw definite conclusions about which factors shaped principals' perceptions. Further research is needed to examine the potential relationship between principals' values and views on education and the type of educational institute where they obtained formal education in school leadership.

For those principals who to a lesser extent perceive test scores as proxies of professional success, social pressure can still contribute to them conforming to the expectations embedded in the tests. This form of pressure can grow when external audiences or society in general come to embrace the definition of education quality promoted by the tests and PBA system, a process which can be reinforced by features of commensuration. On the one hand, this seems to affect low-performing schools in particular, as these schools become increasingly perceived as 'low-quality' schools and sometimes face significant administrative and social pressure to improve their performance, which in turn can reinforce self-fulfilling prophecies. Nonetheless, this study's findings indicate that average- and high-performing schools are not spared from this form of social pressure. With the recent publication of the value-added model, attention has shifted towards whether schools perform in line with what can be expected of them, considering their student body. That is, increasing use of the value-added model by external audiences such as local politicians and the media, following the perception that the value-added model forms a more accurate measure of the school's contribution, implies that high-performing but low-contributing schools are now subject to increased questioning.

The analysis, furthermore, shows that features of commensuration do not only influence external audiences' perceptions of schools, and as such can impact principals' behavior, but also principals' self-perceptions. To assure schools are 'on track', many principals monitor and compare their performance, where previous performance or the performance of other schools or the municipality serve as reference points (see also Skedsmo 2018). This information is then used to reflect on whether there exists a need to adapt organizational, instructional or pedagogical approaches. Similar findings were reported by Feniger et al. (2015, p. 15), who showed that school comparisons based on test scores became 'a major lens through which principals look at their own school and accordingly make decisions'. While some principals interviewed for this study rejected that test scores reflect school quality, the majority of them expressed the belief that test data uncovered an important truth about the school situation and the effectiveness of the schools' teaching practices.

Principals' supportive attitude towards the use of performance data to reflect on and modify educational practices, which aligns to findings reported by Seland et al. (2013), may reflect technological advances, such as improvements made to the test format, as well as development of support systems in how to make productive use of achievement data. At the same time, the relatively broad support can imply a process of naturalization or institutionalization. Whereas initially strongly resisted, the tests seem to have become more accepted over time (see also Gunnulfsen and Møller 2017; Mausethagen 2013), in particular among primary school principals (Seland et al. 2013). Similarly, despite that principals differed in how much importance they ascribe to the achievement of basic skills compared to other educational goals and purposes, all proclaimed to perceive acquisition of basic skills as an important school mission, which resulted in at least a mild interest in how individual students performed at the national tests. Nonetheless, despite principals' willingness to take responsibility for centrally-defined learning goals, and to use achievement data for pedagogical adaptation, most principals remain critical of some of PBA's policy tools, most notably the publication of results. While many argue to value transparency in terms of learning outcomes, the interpretation and use of test scores by external actors, specifically the media and some politicians, is perceived as doing more harm than good.

While the study's findings have provided insight into the reflexive interactions between actors and measures, it is important to highlight that the self-reported nature of the data implies that the findings, in particular with regards to classroom practices, should be regarded as beliefs and intentions, rather than as evidence of what is happening in classrooms (Creswell 2009). Similarly, even though some

principals may have embraced the PBA mandate, they remain for an important part dependent on teachers' willingness and capacity to incorporate real changes in the classroom. Compared to school leaders, teachers have generally positioned themselves more critical to the usefulness of national test data (Seland et al., 2013; Skedsmo 2018), and studies have highlighted how teachers continue to struggle with how to respond to demands from national tests (Gunnulfsen 2017; Mausethagen 2013). Previous research has moreover indicated that willingness to use performance data does not guarantee productive data-use (Gunnulfsen 2017; Mausethagen et al. 2017). This highlights the need for future research on how teachers reflect on and respond to the different mediations of PBA demands employed by principals, possibly combining interviews with teachers with classroom observation.

To sum up, by examining how school principals in different local accountability regimes and at different schools perceive, interpret and translate PBA demands, this study has contributed to opening up the 'black box' of policy enactment in accountability contexts characterized by an ambition to elicit change 'from the inside' by influencing actors' dispositions, as well as a relative absence of material consequences. Moreover, by adopting a mechanism-based approach, and by examining the conditions under which the mechanisms operate, this study has contributed to the understanding of how to interpret the variegated school responses adopted in these accountability contexts. The study's findings highlight how standardized testing and PBA, even in the absence of material consequences and low levels of marketization, can drive behavioral change, by establishing new norms of good educational practice, and by changing how educators make sense of core aspects of their work.

References

- Au, W. (2007). High-stakes testing and curricular control: A qualitative meta-synthesis. *Educational Researcher*, 36(5), 258–267.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215-228.
- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu-net: Globalisation and Education Policy Mobility*. London: Routledge.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How Schools Do Policy: Policy Enactments in Secondary Schools*. London: Routledge.

- Barbana, S., Dumay, X., & Dupriez, V. (2019). Local implementation of accountability instruments in the French-speaking community of Belgium. *European Educational Research Journal*, 19(2), 94–108.
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, 32(4), 585-596.
- Campbell, D. T. (1957). Factors Relevant to the Validity of Experiments in Social Settings. *Psychological Bulletin*, 54(4), 297–312.
- Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*, <https://doi.org/10.1080/02680939.2020.1739337>
- Candido, H. H. D. (2020). Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*, 1-32.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211–244.
- Courtney, S. J. (2014). Post-panopticism and school inspection in England. *British Journal of Sociology of Education*, 37(4), 623-642.
- Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: SAGE Publications, Incorporated.
- Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *The Teachers College Record*, 106(6), 1145–1176.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173-189.
- Espeland, W. N., & Stevens, M. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, 24, 312–343.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Falabella, A. (2014). The performing school: The effects of market and accountability policies. *Education Policy Analysis Archives*, 22(70), 1–26.

- Feniger, Y., Israeli, M., & Yehuda, S. (2015). The power of numbers: the adoption and consequences of national low-stakes standardised tests in Israel. *Globalisation, Societies and Education*, 14(2), 183-202.
- Ferrer-Esteban, G. (2020). *Sampling strategy. REFORMED Methodological Papers No. 3*. Bellaterra: Universitat Autònoma de Barcelona.
- Gunnulfson, A. E., & Møller, J. (2016): National Testing: Gains or Strains? School Leaders' Responses to Policy Demands. *Leadership and Policy in Schools*, 16(3), 455-474.
- Gunnulfson, A. E. (2017). School leaders' and teachers' work with national test results: Lost in translation? *Journal of Educational Change*, 18(4), 495-519.
- Hedström, P., & Swedberg, R. (1998). Social Mechanisms: An Introductory Essay. In P. Hedström & R. Swedberg (Eds.), *Social Mechanisms: An Analytical Approach to Social Theory* (pp.1–31). Cambridge, MA: Cambridge University Press.
- Heffernan, A. (2017). The accountability generation: exploring an emerging leadership paradigm for beginning principals. *Discourse: Studies in the Cultural Politics of Education*, 39(4), 509-520.
- Keddie, A. (2013). Thriving amid the performative demands of the contemporary audit culture: A matter of school context. *Journal of Education Policy*, 28(6), 750-766.
- Larsen, E., Møller, J., & Jensen, R. (2020). Constructions of professionalism and the democratic mandate in education A discourse analysis of Norwegian public policy documents. *Journal of Education Policy*, DOI: 10.1080/02680939.2020.1774807
- Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. In G. Noblit (Ed.), *Oxford Research Encyclopaedia of Education* (pp. 1-28). New York: Oxford University Press.
- Maroy, C. (2015). Comparing Accountability Policy Tools and Rationales: Various Ways, Various Effects? In H. G. Kotthoff & L. Klerides (Eds.), *Governing Educational Spaces: Knowledge, Teaching, and Learning in Transition* (pp. 35-56). Rotterdam: Sense Publication.
- Maroy, C., and Pons, X. (2019). *Accountability Policies in Education. A comparative and multilevel analysis in France and Quebec*. Cham: Springer.
- Mausethagen, S. (2013). Talking about the test. Boundary work in primary school teachers' interactions around national testing of student performance. *Teaching and Teacher Education*, 36, 132-142.
- Mausethagen, S., Prøitz, T., & Skedsmo, G. (2017). Teachers' use of knowledge sources in 'result meetings': thin data and thick data use. *Teachers and Teaching*, 24(1), 37-49.

- Mennicken, A., & Espeland, W. N. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, 45(1), 223–245.
- Merton, R. K. (1968). *Social Theory and Social Structure* (3rd edition). Glencoe, IL: Free Press.
- Mittleman J., & Jennings J. L. (2018). Accountability, Achievement, and Inequality in American Public Schools: A Review of the Literature. In B. Schneider (Ed.), *Handbook of the Sociology of Education in the 21st Century* (pp. 475-492). Cham: Springer.
- Møller, J., & Skedsmo, G. (2013). Modernizing education – NPM reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336–355.
- Piattoeva, N., & Boden, R. (2020). Escaping numbers? The ambiguities of the governance of education through data. *International Studies in Sociology of Education*, 29(1-2), 1-18.
- Pollock, K., & Winton, S. (2016). Juggling multiple accountability systems: How three principals manage these tensions in Ontario, Canada. *Educational Assessment, Evaluation and Accountability*, 28(4), 323-345.
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). Evaluering av nasjonale prøver som system [Evaluation of national testing as a system]. Published online: <https://www.nifu.no/publications/1025043/> Accessed 1 February 2020.
- Sellar, S. (2015). A feel for numbers: Affect, data and education policy. *Critical Studies in Education*, 56(1), 131–146.
- Skedsmo, G., & Mausethagen, S. (2016). Accountability policies and educational leadership – a Norwegian perspective. In J. Easley, & P. Tulowitzki, P. (Eds.), *Accountability and Educational Leadership – country perspectives* (pp. 205-223). London: Routledge.
- Skedsmo, G. (2018). Comparison and Benchmarking as Governing Processes in Norwegian Schools. In L. M. Carvalho, L. Levasseur, M. Liu, R. Normand, & D. A. Oliveira (Eds.), *Education Policies and the Restructuring of the Educational Profession* (pp. 137-158). Singapore: Springer.
- Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731-762.
- Spillane, J. P., & Jennings, N. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *The Teachers College Record*, 98(3), 449–481.

- Statistics Norway. (2019). Pupils in Primary and Lower Secondary School. Statistics Norway. <https://www.ssb.no/en/utdanning/statistikker/utgrs>
- Thiel, C., Schweizer, S., & Bellmann, J. (2017). Rethinking side effects of accountability in education: Insights from a multiple methods study in four German school systems. *Education Policy Analysis Archives*, 25(93), 1–32.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy and Practice*, 21(2), 221-237.
- Verger, A. Parcerisa, L. & Fontdevila, C. (2019). The Growth and Spread of National Assessments and Test-based Accountabilities: A Political Sociology of Global Education Reforms. *Educational Review*, 71(1): 5–30.
- Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020). In and out of the 'pressure cooker': Schools' varying responses to accountability and datafication. In S. Grek, C. Maroy & A. Verger (Eds.), *World Yearbook of Education 2021*. London: Routledge

CHAPTER 5: SUMMARY OF MAIN RESEARCH FINDINGS

The following overarching research question guided the research reported on in this dissertation: How is performance-based accountability, as a global policy for education reform, adopted, shaped and enacted in the Norwegian education system? In order to shed light on this overarching question, three sub-studies were conducted, which focus on different scales in the deployment of accountability policies. Each sub-study was guided by a specific research question. This chapter provides an overview of the main findings of this dissertation, thereby formulating answers to the research questions that guided the study.

Chapter 2 focused on the following research question: “Why did Norway adopt PBA in education at the turn of the millennium, and how did central authorities operationalize and develop PBA through different policy instruments and tools over a period of 15 years (2003-2018)?” The findings reported on in Chapter 2 reveal that a ‘window of opportunity’ for major reform (cf. Kingdon, 1995) opened shortly after the turn of the millennium, when the arrival of a Conservative government coalition coincided with the publication of the first PISA results. Both government officials and media outlets scandalized the below-expected PISA results (see also Elstad, 2012), which contributed to the perception of a ‘learning crisis’ in Norwegian education. To an important extent, this perceived crisis was attributed to the ways Norwegian schools had been governed.

The chapter furthermore outlines how, in order to modernize and raise the performance and equity of the education system, a shift in school governance was proposed. Centralized steering by means of input-regulations was to be replaced by an outcome-based curriculum, in combination with a further devolution of pedagogical and organizational decision-making power to local education authorities and schools. The establishment of a National Quality Assessment System, and in particular the adoption of census-based, large-scale assessments, were suggested as key steering devices to monitor and control whether student achievement was in line with centrally-defined objectives. Moreover, the publication of test results, which would enhance visibility, comparison and benchmarking of school and municipal performance, was seen as a context-appropriate pressure mechanism to ensure key actors would be responsive to and take responsibility for learning outcomes. While in the aftermath of the ‘PISA shock’, the introduction of national tests was considered necessary by an increasing number of key actors, great controversy remained to exist regarding the establishment of a web portal where test results would be published. The chapter highlights how both the framing of the policy

proposal, as well as the decision-making process, enabled the relatively abrupt adoption of both measures in 2003.

With regard to the key drivers behind the institutional transformation in Norway, the chapter reveals how the perceived importance of educational achievement in basic skills such as numeracy and literacy, with an eye on future economic potential and competitiveness, contributed to the shock the PISA results elicited, as well as the perceived necessity to rapidly raise learning outcomes. Simultaneously, Chapter 2 highlights how the adoption of PBA in Norwegian education cannot be seen in isolation from broader changes in public sector governance, inspired by NPM policy principles. In important ways, the adoption and institutionalization of PBA was guided by a belief in the need to steer the public sector by means of outcome measures, transparency, comparison and accountability, in order to raise its efficiency and effectiveness. In this regard, international expert advice and research documents seem to have served both as a source of inspiration and a source of justification during policy design processes. Nonetheless, despite a clear global imprint on recent educational reforms in Norway, the country's political institutions and social-democratic welfare regime have mediated the operationalization of PBA, as well as its institutionalization within existing regulations and practices. In this light, Chapter 2 highlights how a final driver behind the spread of PBA to Norway relates to accountability being an example of an 'empty vessel' policy, which can be adopted to serve a diverse set of goals (Steiner-Khamsi, 2016). Unlike early adopters, where PBA often served to promote market-based reforms (Verger, Parcerisa, & Fontdevila, 2019), Norway's adoption of PBA served to ensure equity and quality standards in a highly decentralized education system.

More specifically, with regard to the question how central authorities operationalized and developed PBA through different policy instruments and tools over a period of 15 years (2003-2018), Chapter 2 highlights how administrative control mechanisms, which monitor, control and make visible school and municipal performance, have been layered on top of one another over time. Nonetheless, whereas increased visibility, benchmarking, and administrative control form key mechanisms in putting pressure on local and school actors to ensure they adapt their practices in line with the competency aims formulated in the national curriculum, at least at the central level, the system remains characterized by limited use of external sanctions or incentives. Moreover, beyond external control mechanisms, the system relies on internal measures designed to mobilize actors' internal feelings of responsibility and self-reflection.

A final finding of Chapter 2 is that, even though central policy discourse remains characterized by a belief in the professional responsibility of local and school actors to use performance data to foster school improvement, interpretations differ at local scales. That is, it appears that various municipal superintendents have decided to enact quality assessment systems based on detailed and in advance specified performance indicators, risk assessment, publication of school test scores and performance contracts.

Building on the finding that within Norway's PBA system, the publication of test scores forms a key external pressure mechanism that serves to elicit the desired change in local and school actors' behavior, Chapter 3 examined the role of the media in mediating the impact of this pressure mechanism. More specifically, this chapter focused on the following research question: "What are the most recurrent discursive frames used by the Norwegian regional and local press when informing readers about national standardized tests, and to what extent does the pervasiveness of particular frames differ over time and across geographical localities?" The findings reported on in Chapter 3 highlight how the Norwegian regional and local press, ever since the introduction of the tests in 2004, have reported on national standardized testing on a yearly basis. The largest peak in press coverage was found in 2005, one year after the tests had been introduced. Ever since 2007, press coverage has remained relatively stable, including during those years (2007-2014) that test results were not published at the school level. In this regard, the analysis reveals that even though school results remained hidden, media outlets found ways to get hold of school level data, resulting in 619 newspaper articles that presented school results in the form of comparisons and/or rankings during 2007-2014. Moreover, by relying on framing theory (Entman, 1993), the analysis reported on in Chapter 3 identifies four dominant frames, present in 2,475 newspaper articles (i.e. 81.25% of the analyzed corpus). The four frames, which were named 'performance', 'transparency and empowerment', 'misinterpretation and misuse' and 'criticism', communicate distinct problem definitions, causal interpretations, moral evaluations, and treatment recommendations (cf. Entman, 1993). The analysis, furthermore, shows how the dominance of each of the four frames varies over time as well as across geographical localities.

More specifically, the analysis reveals how during the first two test rounds, the frames of 'misinterpretation and misuse' and 'criticism' were particularly pervasive. Both frames communicate a critical stance towards the testing system, although the frames differ in important regards. Whereas the frame of 'criticism' rejects the validity, reliability and usefulness of the national tests, the frame of 'misinterpretation and misuse' suggests that the test results can have important pedagogical and

organizational value, if correctly interpreted and used. In this regard, the frame of ‘misinterpretation and misuse’ is critical towards the publication of results, which is said to elicit a situation in which results become a goal in itself. It is perhaps little surprising that these two frames dominated during 2004 and 2005, as the first two test rounds were met by severe societal discontent, student boycotts and destructive evaluation reports (Lie et al., 2005). Upon reintroduction of the tests in 2007, following quality improvement efforts and a revision of the conditions of the tests’ administration (as outlined in Chapter 2), both frames decline in importance.

Consequently, ever since 2007, the frame of ‘performance’ has become the most dominant frame in the articles analyzed, followed by the frame of ‘transparency and empowerment’. The latter, similar to the frames of ‘misinterpretation and misuse’ and ‘criticism’, communicates an explicit opinion on the validity, reliability and usefulness of the national test system. That is, the frame of ‘transparency and empowerment’ promotes the idea that national tests are valid, reliable and useful instruments to uncover students’ learning difficulties, thereby providing key knowledge that empowers educators and authorities to adapt education to individual and local needs and address social and geographical inequalities. The frame, moreover, strongly criticizes efforts to discontinue the publication of school results, presenting reduced transparency as a direct threat to inclusion efforts. The use of arguments relying on social democratic principles such as inclusion and social equality is interesting, and might be a particularly effective way of convincing readers of the usefulness and importance of national testing in a country like Norway, where for decades, the comprehensive school model has sought equal opportunity for all, regardless of gender, socioeconomic or ethnic background, or geographical location (Blossing, Imsen, & Moos, 2014).

In contrast, the frame of ‘performance’ largely refrains from debating the validity, reliability, fairness or usefulness of the national tests. Rather, this frame uses test scores to report on, compare, praise and blame schools, municipalities and counties for performance. In so doing, the frame communicates a range of implicit messages, including that test performance reflects the effort put in by key actors, that hard work results in high performance, and that low performers should take urgent action to assure improvement of learning outcomes. While such newspaper articles appear not to contain any explicit opinion or reflection on the tests, they in fact only appear neutral. That is, by presenting test results as non-disputed facts, the message is conveyed that the tests are a valid measure of educational quality and that schools located in different contexts, with different student bodies, histories and values, can be compared. Notions such as ‘the best schools’ only serve to reinforce this message. The

conceptual metaphor of national testing as a competitive sport also conveys a range of implicit messages that enhance school competition and diffuse the idea that test results reflect merit.

The geographical analysis confirms that the frame of ‘performance’ does not only form the dominant frame over most of the time-span analyzed, but also in 15 out of 19 Norwegian counties. Only in three other counties (i.e. Hedmark, Nord-Trøndelag and Oslo), another frame was found more dominant. That is, in both Hedmark and Nord-Trøndelag, the frame of ‘transparency and empowerment’ was found most pervasive, whether in Oslo, the frame of ‘criticism’ was most dominant. Finally, in one county, Hordaland, two frames were found equally pervasive, namely the frame of ‘performance’ and the frame of ‘criticism’.

Chapter 3 ends with a discussion of the possible reasons behind the dominance of the frame of ‘performance’. It is suggested that the predominance of this frame might reflect an increasingly widespread acceptance of national tests as valid and useful measures and crucial indicators of education quality. In reporting on test performance, the media may consider that they fulfil an important role as ‘watchdogs’ for the benefit of the society (Finkelstein & Ricketson, 2012), drawing attention to quality differences and holding ‘responsible’ actors accountable. Simultaneously, the increasing acceptance of the tests might have led to a less explicit debate in the media on the tests, and a decline in importance of the other three frames. Nonetheless, the dominance of the frame of ‘performance’ could also reflect the seductions of quantification (Espeland & Stevens, 2008; Merry, 2016). That is, indicators such as test scores appear to provide objective, scientific, concrete and politically neutral knowledge of a complex world, thereby appealing to ‘the desire for simple, accessible knowledge and to a basic human tendency to see the world in terms of hierarchies of reputation and status’ (Merry, 2016, p. 1). As such, test scores form particularly attractive content for journalists. Finally, it could be the case that comparisons between entities based on averages, as well as over time, correspond to the easiest and cheapest way for non-technical stakeholders such as journalists, who may not have the expertise to fully understand the complexity of the results (Wu, 2010), or the factors behind them, to make sense of and report on national tests.

The analysis conducted in Chapter 3 confirms that, in addition to variability in local accountability regimes (as highlighted by previous research), important geographical differences exist in Norway in how the regional and local press report on and frame standardized testing. As such, the role played by the local press as ‘third-party account-holders’ (Busuioc & Lodge, 2016) differs in important regards in different local contexts. With the aim of gaining a deeper understanding of how such local variability

mediates how school actors experience, make sense of and respond to PBA demands, the sub-study reported on in Chapter 4 was conducted in municipalities which differed according to (a) the level of 'strictness' of the local accountability regime; and (b) the level of involvement of 'third-party' account-holders (i.e. parents and the media). More specifically, Chapter 4 examined the following research question: "How do Norwegian school principals reflect on and respond to PBA demands, and how can variation in principals' response patterns be explained?" In line with previous research (e.g. Barbana, Dumay; & Dupriez, 2019; Candido, 2020; Diamond & Spillane, 2004; Maroy & Pons, 2019; Verger, Ferrer-Esteban, & Parcerisa, 2020), the findings reported on in this chapter highlight important variation in how principals perceive, interpret and respond to accountability demands. Specifically, the analysis identifies three distinct response patterns; i.e. alignment, balancing multiple purposes and symbolic responses.

The first response pattern is adopted by principals who strongly align to the PBA mandate and action theory, and attempt to meet accountability expectations by adopting top-down, performance-oriented management practices and by relying on data-driven decision-making to plan school improvement. The perceived importance of making sure their students achieve basic skills - so to give them a chance to get ahead in life - seems in particular at the lowest performing schools to contribute to a degradation of other identities and organizational purposes. The second response pattern is articulated by principals who only partially agree with the PBA mandate and action theory. While they acknowledge the importance of cognitive skills, they remain wary of a sole emphasis on academic achievement. In other words, they continue to balance multiple educational purposes in their development projects and focus areas. In a similar vein, while they support the expectation to use achievement data to foster school improvement, they do so on their own terms. That is, in comparison with principals in the first response category, decision-making at schools in the second category seems to a lesser extent steered by achievement data. Rather, test results form one out of various sources of information, which foster collective reflection and decision-making. Finally, principals in the third response category express a highly critical attitude towards the PBA mandate and action theory. They question the validity and usefulness of the national tests, and are particularly worried about how the debate on education quality increasingly focuses on how schools perform at national tests. In their development projects and focus areas, they strongly emphasize a focus on the non-measurable. Moreover, they respond predominantly symbolically to the expectation to use achievement data to foster school improvement.

Beyond identifying different response patterns in how principals perceive, interpret and respond to PBA demands, a key objective of Chapter 4 is to explain variation in response patterns, and to determine the conditions under which particular responses are more likely to be articulated. That is, by adopting a mechanism-based analytical approach, and by establishing the conditions under which the mechanisms operate, the chapter aims at offering explanations for ‘*why* we observe what we observe’ (Hedström & Svedberg, 1998, p. 9). More specifically, the chapter relies on two social mechanisms, identified by Espeland and Sauder (2007, p. 6), which induce reactivity, understood as the way ‘individuals alter their behavior in reaction to being evaluated, observed, or measured’.

In so doing, Chapter 4 shows how principals in the first response category, who strongly align to the PBA mandate and action theory, have to an important extent internalized the conception of education as embedded in standardized testing and PBA. Their actions, in turn, increasingly serve to become like the definition of “a good school” promoted by the accountability system. This response pattern is an example of how standardized testing and PBA, by imposing a particular definition of education and encouraging actors to see themselves and behave according to the norms of good practice, can create self-fulfilling prophecies (cf. Espeland and Sauder 2007). The analysis, moreover, reveals how a number of conditions seem to favor principals’ internalization of performance metrics, which relate both to principals’ professional trajectories and the local accountability regime.

With regard to the second response pattern, employed by principals who only partially agree with the definition of “a good school” embedded in the PBA system, the analysis highlights how the new social relations constructed by commensuration form a more central explanatory factor to understand this pattern. That is, the analysis reveals that the terms under which principals assess ‘progress’ and assure their school is ‘on track’ are affected by the new, precise and hierarchical relationships between schools that commensuration creates. Increasingly, comparisons of the school’s performance to previous performance, as well as to performance of other schools or the municipal average, shape how principals perceive their schools and accordingly make decisions (see also Feniger, Israeli, & Yehuda, 2015). On the one hand, this response pattern is promoted by principals’ belief in the validity of the test, which fosters their willingness to use achievement data for school improvement purposes. On the other hand, it seems that the construction of performance rankings by some municipal superintendents or media outlets can provide an additional push to adapt educational practices, so to avoid public humiliation.

The final response pattern is employed by principals who experience a significant mismatch between their views on education and the PBA mandate and action theory. Despite that all principals in this response category claimed to experience that more and more schools are evaluated, compared and judged by means of standardized test scores, most of them were able to keep running their schools as they saw fit. This meant that they, in line with policy expectations, had established systems to analyze test scores, but that this remained a rather isolated exercise, with a limited impact on school organization and educational practices. Nonetheless, under certain conditions, principals felt forced to respond in ways that conflicted with their own principle beliefs and values. In these cases, significant social pressure was exercised, either by local politicians, the media or parents, to aim at higher test performance. The analysis highlights how this form of pressure can grow when external audiences or society in general come to embrace the definition of education quality promoted by the tests and PBA system, a process which can be reinforced by features of commensuration. In turn, social pressure to improve performance can contribute to school actors conforming to the expectations embedded in the test system thereby generating self-fulfilling prophecies (cf. Espeland & Sauder, 2007).

To sum up, with an eye on the overarching research question, the findings of this dissertation highlight how PBA has been adopted and institutionalized in Norway as a way to secure equity and quality standards in a highly decentralized education system. While global drivers and policy ideas have clearly influenced the selection of policy instruments, the trajectory of PBA in Norway is simultaneously shaped by the country's political institutions, welfare regime, and education system, as well as by policy enactment processes. In ensuring educators are responsive to and take responsibility for learning outcomes, the system relies on administrative control, visibility and benchmarking as key pressure mechanisms. At the same time, local authorities, as well as regional and local media outlets, play a key role in shaping how central policy demands play out in practice, thereby influencing policy enactment processes at the school level. In part as a result of significant local variability in how PBA demands are mediated at regional and local scales, school principals' responses to accountability demands differ in important regards. Whereas some principals attempt to accommodate accountability expectations, others negotiate their own terms, or respond predominantly symbolically to the expectation to raise test scores and use achievement data for school improvement purposes. The findings hereby highlight that also in Norway, variation in policy responses and outcomes seems to be the rule, rather than the exception.

References

- Barbana, S., Dumay, X., & Dupriez, V. (2019). Local implementation of accountability instruments in the French-speaking community of Belgium. *European Educational Research Journal, 19*(2), 94–108.
- Blossing, U., Imsen, G., & Moos, L. (2014). *The Nordic model. A 'School For All' Encounters Neo-Liberal Policy, Policy Implications of Research in Education* (Vol. 1). Dordrecht: Springer Series.
- Busuioc, E.M., & Lodge, M. (2016). The reputational basis of public accountability. *Governance, 29*(2), 247-263.
- Candido, H. H. D. (2020). Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education, 29*(1–2), 1–32.
- Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *The Teachers College Record, 106*(6), 1145–1176.
- Elstad, E. (2012). PISA Debates and Blame Management among the Norwegian Educational Authorities: Press Coverage and Debate Intensity in the Newspapers. *Problems of Education in the 21st Century, 48*, 10–22.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology, 113*(1), 1–40.
- Espeland, W. N., & Stevens, M. L. (2008). A Sociology of Quantification. *European Journal of Sociology, 49*(3), 401-436.
- Feniger, Y., Israeli, M., & Yehuda, S. (2015). The power of numbers: the adoption and consequences of national low-stakes standardised tests in Israel. *Globalisation, Societies and Education, 14*(2), 183-202.
- Finkelstein, R., & Ricketson, M. (2012). *Report of the independent inquiry into the media and media regulation*. Canberra, AU: Department of Broadband, Communications and the Digital Economy.
- Hedström, P., & Swedberg, R. (1998). Social Mechanisms: An Introductory Essay. In P. Hedström & R. Swedberg (Eds.), *Social Mechanisms: An Analytical Approach to Social Theory* (pp. 1–31). Cambridge, MA: Cambridge University Press.
- Kingdon, J. W. (1995). *Agendas, alternatives, and public policies*. New York: HarperCollins College Publishers.

- Lie, S., Hopfenbeck, T. N., Ibsen, E., & Turmo, A. (2005). *Nasjonale prøver på ny prøve [National tests retested]*. Oslo: Department of Teacher Education and School Research, University of Oslo.
- Merry, S. E. (2016). *The seductions of quantification: measuring human rights, gender violence and sex trafficking*. Chicago and London: University of Chicago Press.
- Maroy, C., and Pons, X. (2019). *Accountability Policies in Education. A comparative and multilevel analysis in France and Quebec*. Cham: Springer.
- Steiner-Khamsi, G. (2016). New directions in policy borrowing research. *Asia Pacific Education Review*, 17, 381–390
- Verger, A. Parcerisa, L. & Fontdevila, C. (2019). The Growth and Spread of National Assessments and Test-based Accountabilities: A Political Sociology of Global Education Reforms. *Educational Review*, 71(1), 5–30.
- Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020). In and out of the 'pressure cooker': Schools' varying responses to accountability and datafication. In S. Grek, C. Maroy & A. Verger (Eds.), *World Yearbook of Education 2021*. London: Routledge.
- Wu, M. (2010). Measurement, Sampling, and Equating Errors in Large-Scale Assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.

CHAPTER 6: CONCLUSIONS

Over the past decades, an increasing number of countries have adopted or strengthened their PBA systems in education (Ball, Junemann, & Santori, 2017; Verger, Parcerisa, & Fontdevila, 2019), despite a lack of strong evidence regarding the effectiveness of PBA in raising student achievement, fostering instructional improvement, or reducing educational inequalities (e.g. see Falabella, 2014; Jarke & Breiter, 2020; Maroy, 2015; Maroy & Pons, 2019; Sahlberg, 2016; Verger & Parcerisa, 2017). While the global popularity of PBA has sparked significant research interest, the existing educational accountability literature remains characterized by important gaps, which limit our understanding of why policymakers, in such diverse contexts, turn towards PBA to reform their education systems, as well as the mechanisms and conditions under which PBA produces particular policy outcomes.

This dissertation contributes to filling some of the gaps in the existing literature by adopting a multi-scalar case-study, which scrutinizes the different (but interconnected) scales in the deployment of accountability policies in Norway. The multi-scalar approach translated into three sub-studies, which illuminate the deployment of accountability policies at different scales, while continuing to shed light on how policy trajectories, enactment processes, and contextual contingencies interact in complex and multiple ways. In this concluding chapter, the methodological strengths of the different sub-studies are briefly highlighted. Thereafter, the epistemological value of a multi-scalar approach when examining the deployment of accountability policies is discussed. Subsequently, the empirical and theoretical contributions of this dissertation are highlighted, followed by a discussion of the limitations of the study, and ideas for future research. This concluding chapter ends with a discussion of the importance of the results from a policy point of view.

Methodological insights

With regard to the methodological contribution of this dissertation, it is important to emphasize the specific methodological strengths of the sub-studies. More specifically, the first sub-study highlights the value of expert interviews to understand the often implicit or unarticulated ‘world views’ informing policy tool selection (Maroy, 2015). Moreover, by focusing not only on the moment of policy adoption, but by tracing policy evolution over a period of 15 years, this sub-study provides a more comprehensive view on how PBA has been operationalized and institutionalized in Norwegian education over time. In addition, the second sub-study confirms the need for systematic analyses of

media coverage, so to identify potential variation in media coverage over time and across geographical localities. Indeed, the results highlight that in Norway important geographical differences exist in how regional and local media outlets report on and frame standardized testing. Variation in media coverage within the same country may explain in important ways the performative pressure that school actors experience, which can generate far-reaching effects on school practices and responses to PBA reforms. Geographical variation in media coverage therefore deserves a critical spotlight. Finally, the sampling approach of the third sub-study forms an important methodological strength. Too often, studies on the deployment of accountability policies rely on convenience sampling techniques, thereby neglecting important theoretically relevant criteria that could affect the phenomenon under study. By relying on a two-phase sampling procedure, and by including theoretically relevant sampling criteria, the third sub-study strengthened the validity of the results.

The value of a multi-scalar approach

This dissertation highlights the value of a multi-scalar approach to understand the deployment of accountability policies. The different sub-studies included in this dissertation provide evidence for how both policy trajectories and enactment processes are shaped by actors, processes and mechanisms operating at different scales. More specifically, the first sub-study highlights how the orientation and meaning of the PBA trajectory is shaped both by global policy discourses as well as processes of policy enactment at intermediate scales and at the school level. For example, significant local variation in how municipal superintendents took on their role as formal account-holders, as well as the perceived passive attitude of parents towards performance metrics, seem to have guided recent decisions to raise pressure on local authorities. In a similar vein, the contentious origins of PBA instruments in Norway are related to the critical responses that these instruments generate among key stakeholders.

Moreover, the second and third sub-study highlight how central policy demands undergo significant negotiation, co-construction and translation before they finally reach schools. In this light, this dissertation points out how policy instruments can start to live a life of their own, ‘independent of the decisions that created them’ (Kassim & Le Galès, 2010, p. 11), thereby taking forms and producing outcomes that may contradict or extend beyond initial policy goals (Le Galès, 2010). Indeed, it appears that a number of municipal superintendents, taking advantage of accountability reforms at the national level, undertake efforts to regulate school actors’ behavior by building on closely aligned policy tools,

including in advance specified performance indicators, risk assessment, and performance contracts. These efforts seem to be guided by the assumption that many aspects of teaching and learning can be controlled and documented, and that holding school leaders and teachers accountable for students' results forms an efficient and effective way to raise performance. In a similar vein, central authorities have not been able to hamper the construction of performance rankings by media outlets, despite various attempts. As shown by the second sub-study, even during the years that school results were not published at the 'School Portal', media outlets got access to school data and continued to report on performance differences, although coverage is not homogenous across regions.

These local differences highlight the importance of examining how actors operating at different scales act as 'brokers' of predefined policy goals when making sense of school responses to accountability demands (see also Proitz, Mausestagen, & Skedsmo, 2019). Indeed, the third sub-study shows how the ways in which school principals reflect on and respond to PBA demands is in important ways shaped by the local accountability regime and the level of activity of 'third-party' accountholders, in particular the local media. With regards to the latter, principals' accounts revealed that an active local press, which uses test scores to compare, blame and praise schools and educators, can contribute to the public perception of test scores as relevant and robust indicators of school quality, thereby raising pressure on schools to aim at high test performance, so to avoid public humiliation. This finding aligns to previous research, which shows that active media coverage of test scores can generate parental interest in test results (Bjordal, 2016), while contributing to pressure felt by Norwegian educators, sometimes leading to the adoption of 'panic measures' with potential harmful effects (Elstad, 2009).

As such, beyond demonstrating how policy design, policy enactment and contextual contingencies interact in multiple and complex ways, this dissertation shows how, without a comprehensive understanding of how different scales of governance or public action interact, it may remain difficult to fully understand and explain outcomes at a particular scale of interest. At the same time, by focusing on different scales of public action when examining the deployment of accountability policies, the risk of reductionism or bias can be lessened (Maroy & Pons, 2019). That is to say, a sole focus on accountability policies at the central level might have led to the conclusion that PBA in Norway remains characterized by significant trust in the profession and little use is made of external incentives to instigate behavioral change. By including the regional and local scales in the analysis, it becomes clear that this conclusion would need to be nuanced. Namely, it appears that a number of municipal superintendents increasingly monitor and control school performance, while raising the stakes

attached to learning assessments. Simultaneously, in a number of contexts, regional and local media outlets further raise pressure, by reporting to the public on performance indicators as proxies of school quality.

In similar vein, a bias could emerge when studies on the deployment of accountability policies focus on a number of schools, while failing to sufficiently contextualize policy trajectories and the local policy context. As a result, conclusions could be drawn about how school leaders or teachers experience and respond to accountability policies, without paying attention to how such experiences and responses have been shaped by the local institutional and social context (Maroy & Pons, 2019).

The above highlights the need to include different scales in an analysis of the deployment of accountability policies – but also other types of educational reforms. On the one hand, an understanding of why global policy discourses such as PBA are adopted, and how these policies are operationalized and institutionalized in specific settings forms a first step towards understanding the policy outcomes of PBA more comprehensively. On the other hand, an understanding of how central policy demands are interpreted, translated and co-constructed at different scales might provide a more complete understanding of policy trajectories.

Accountability in a Nordic education setting: Learning from the Norwegian experience

This dissertation furthermore provides novel empirical evidence on a lesser documented accountability regime in a Nordic educational setting. In this section, a reflection is added on what we can learn from the Norwegian experience.

With regard to our understanding of the drivers behind the global advance of PBA, the Norwegian case highlights how the rationales for adopting PBA are highly context-specific. Rather than to promote market-based reforms, PBA has been adopted and institutionalized in Norway to secure equity and quality standards in a highly decentralized education system. This highlights how the understanding of accountability as ‘an empty vessel’ policy, which can be adopted to serve a diverse set of goals (Steiner-Khamsi 2016), is key in order to make sense of the global spread of PBA, including to contexts formerly considered ‘immune’ to this globalizing policy phenomenon (Verger, Parcerisa, & Fontdevila, 2019).

Moreover, the Norwegian experience adds evidence to the observation that we are far from witnessing global convergence in education systems or around a specific model of PBA. Rather, the Norwegian example illuminates the crucial role played by a country's history, political institutions, welfare regime, and education system in mediating the trajectory of PBA policies. On the one hand, global drivers and policy trends have had a clear influence on policy design processes, both in contributing to a perceived need to revisit existing policies, as well as in inspiring and legitimating policy solutions. On the other hand, Chapter 2 highlights how global policy ideas are strongly filtered, negotiated and adapted by actors operating at different scales. For example, some elements of this global reform package, including a promotion of market competition, or the use of material incentives such as performance-based pay, have been left out (at least at the central level), as they were not perceived as contextually appropriate. As a result, rather than a neoliberal variant of PBA, the Norwegian way of PBA reflects a 'quality assurance' approach, which emphasizes the vertical and bureaucratic regulation of schools. By means of various administrative regulatory instruments, the actions of school actors and authorities are attempted to be regulated and controlled.

As such, the Norwegian case can shed light on policy outcomes associated with PBA in a context that diverts in important regards from the over-researched early-adopters, where 'hard' or 'strict' accountability approaches have been developed. As highlighted by Chapter 4, no homogenous effects on school organization and pedagogical practices are found. Rather, principals' response patterns differ in important ways. Nonetheless, the majority of the interviewed principals disagreed with a predominant focus on academic achievement, and continued to place emphasis on the broad purposes of education. In a similar vein, even though that at some schools, test results increasingly inform decision-making processes, many principals continue to rely on teachers' professional judgement and collective decision-making processes when introducing pedagogical adaptation.

At the same time, in various local contexts, characterized by a strong performance orientation, and a long history of PBA demands, some of the perverse effects associated with 'hard' or 'strict' accountability systems seem to be appearing. This relates in particular to the strong emphasis on tested content at the expense of non-tested skills and competencies. That is, Chapter 4 highlights how a number of schools are increasingly guided by a managerial logic of efficacy, characterized by a strong preoccupation with school performance, and extensive use of different mechanisms that serve to "manage" pedagogical practices. At these schools, student performance is increasingly tracked, monitored and analyzed, teachers are asked to justify and explain performance, and pedagogical

adaptation is steered by “what works” and “evidence-based practices” so to raise test results. In important ways, this preoccupation with performance redefines the conception of “a good school” and “a good teacher”. At the lowest performing schools, the interview data indicated a tendency towards degradation of other identities and organizational purposes, including a formerly strong emphasis on relevance for students’ multicultural identity. While this dissertation was only able to capture the impact on classroom practices in indirect ways and the findings with regard to classroom practices should therefore be interpreted with caution (see section ‘limitations and future lines of inquiry’), the interview data highlighted how the focus of some schools seems to become increasingly narrow and determined by performance metrics, thereby contributing to curricular reduction.

Understanding the reactive interactions between actors and measures

To understand and explain variation in policy responses, Chapter 4 adopted a mechanism-based explanatory approach. In so doing, this dissertation moves beyond the observation *that* a gap exists between policy expectations and policy outcomes, and towards an explanation of *why* this gap exists and comes about. In this light, a key theoretical contribution of this dissertation is that PBA can drive behavioral change by reframing norms of good educational practice, and by changing how educators make sense of core aspects of their work. This finding challenges a central premise of educational accountability research; i.e. that external incentives and sanctions, often of a material nature, form the prime explanatory factors of behavioral change instigated by PBA reforms. Rather, this dissertation’s findings highlight the importance of how principals make sense of performance metrics, and how they interpret what such metrics tell them about themselves and their schools. Moreover, the findings highlight multiple determinants of performative pressure, beyond pressure exercised by the formal account-holder or objective accountability system. On the one hand, it appears that some principals increasingly impose pressure on themselves to obtain high test scores. Rather than being motivated by self-interest or external incentives, self-imposed pressure seems to derive from a genuine belief in the need to raise test performance so to give students a chance to get ahead in life. On the other hand, it appears that some principals experience significant social pressure to aim at high test scores, exercised by ‘third-party’ account-holders such as parents and the media. As such, by adopting a mechanism-based explanation inspired by literature in the field of sociology of quantification, and by establishing the conditions under which the mechanisms operate, this dissertation contributes to

advancing the understanding of why and under what circumstances PBA is producing particular policy responses.

Limitations and future lines of inquiry

The research reported on in this dissertation is not free of limitations. Nonetheless, when reflecting on the shortcomings of this research, fruitful directions for future research appear. More specifically, this study has not generated primary data on how local authorities interpret, translate and co-construct central accountability demands, and future research could address this gap. By means of primary data, generated for example through in-depth interviews with local authorities, a deeper understanding could be gained of how local authorities negotiate and translate policy demands, thereby shaping how accountability plays out in practice. Considering the data collection demands of the three sub-studies, it did not prove feasible to also conduct in-depth interviews with local authorities. As such, to develop an understanding of differences in local accountability regimes, this study relied on (a) policy documents; (b) previous research; and (c) quantitative data derived from the administration of an electronic questionnaire. While these sources at least formed a partial way to overcome this limitation, future research would benefit from an in-depth examination of how the ways in which local authorities make sense of central policy demands shapes policy enactment processes at the school level.

The second sub-study documents important geographical differences in media coverage, but does not explain why these differences exist. Potential determinants of geographical differences could relate to local media landscapes (e.g., the political ideology of local newspapers, newspapers' resources), local differences in performance, local policy contexts, or other cultural specificities. While an in-depth examination of the determinants behind geographical differences lies beyond the scope of this sub-study, they could be explored by future research.

The third sub-study relies on self-reported data by school actors. This implies that the findings reported on in Chapter 4 regarding the impact of PBA on pedagogical practices should be interpreted as beliefs about actions, rather than as evidence of what is effectively happening in classrooms. A possible way to overcome this limitation in future research entails conducting classroom observation. Moreover, various studies have demonstrated the value of observing team meetings (either in person or by recording them), in order to develop a better sense of how test data and accountability demands are responded to in practice (e.g. see Gunnulfsen, 2017; Mausethagen, Proitz, & Skedsmo, 2017).

Furthermore, the third sub-study solely focuses on principals, reason why it does not claim to be capturing all possible responses to PBA at the school level. Principals were considered a particularly interesting group of school actors as they are primarily held responsible for school results. Nonetheless, they depend in important ways on teachers to put policy demands into practice. In particular since Norwegian teachers have tended to position themselves as more critical towards national test data (e.g. Mausethagen, 2013; Gunnulfsen, 2017), it is crucial to examine how teachers reflect on and respond to PBA demands in different schools and local policy contexts, so to be able to advance our understanding of how PBA affects pedagogical practices and social relations within schools.

Finally, it is important to emphasize the urban focus of the third sub-study. With over 82% of the Norwegian population residing in urban areas (Statistics Norway, 2019), many municipalities and schools in rural areas are small. Moreover, important differences exist between urban and rural municipalities with respect to the role played by local authorities as formal account-holders (Seland et al., 2013). Consequently, there might be reason to suspect that different findings would have been obtained had this study been conducted in a selection of rural municipalities. At the same time, recent evidence suggests that even in more remote municipalities, where municipal superintendents play a limited role in prescribing measures and following-up schools for performance and school competition is absent, some teachers explain to experience performative pressure (e.g. see Haugen, 2018; Theodorsen, 2017). As argued by Skedsmo (2018), in rural municipalities, the stakes might be higher or more personal for individual school leaders and teachers due to small and transparent communities. This highlights the need for an examination of the multiple determinants of performative pressure, and to include both urban and rural municipalities in such a study.

In the same way that reflecting about the limitations of this study is informative for future research directions, its main findings also derive into ideas about future lines of inquiry. For example, the finding that a substantial number of newspaper articles rely on test scores to praise educators or schools for test performance generates a number of questions. First of all, a deeper examination of the images conveyed by newspaper articles about what constitutes a ‘good teacher’ and a ‘successful school’ is needed. This is crucial, as praising educators for test performance might appear neutral or harmless, while at the same time, by means of such practices, particular ideas of what counts as successful schools and good teachers may be created or reinforced. Such ideas could potentially have far-reaching effects, including an influence on ‘how people think about what they do, what is

comparable, how excellence or mediocrity is defined, and even who we are' (Espeland & Sauder, 2012, p. 86).

Second of all, there exists a need for more research on the impact of different forms of media coverage. To date, most studies have focused on the responses by educators in blamed schools (Elstad, 2009), while less remains known about how other media practices, such as praising, rankings and comparisons, affect different schools, including those ignored by the media. Such research might be particularly relevant in contexts where schools compete and test scores affect the school's reputation. For example, whereas praising may seem like harmless media practice, in small municipalities where only a few schools are located, praising one school de facto means neglecting or even silently blaming the other few schools. Measuring the effect of media coverage is, however, a complex task, as it is difficult to determine the resonance of each article. Indeed, while Chapter 3, in contrast to previous research (Baroutsis & Lingard, 2017; Elstad, 2009), found blaming to be a less dominant media practice than praising, the impact of one single harsh article in a small municipality may be more important than several articles reporting scores without praising or blaming any school in particular.

Policy relevance

The observation that PBA policies have the potential of producing a number of perverse or undesired effects is neither new, nor recent. For over two decades, scholars have documented how PBA policies, in particular when high stakes are attached to them, can generate effects such as teaching to the test, curriculum narrowing, educational triage and cheating (e.g. Au, 2007; Booher-Jennings, 2005; Jacob & Levitt, 2003; Jennings, 2010; Reback, 2008). In recent years, policymakers in different contexts have recognized this, and have made adjustments in order to prevent perverse effects from occurring. For example, in both Germany and Israel, PBA systems were deliberately designed without attaching high stakes (Feniger, Israeli, & Yehuda, 2015; Thiel, Schweizer, & Bellmann, 2017). In Norway too, policymakers were aware of perverse effects in contexts such as the UK and the US. To paraphrase one policymaker interviewed for the first sub-study, precisely because they work (but not always in the desired ways), material incentives such as performance-based are not considered in the Norwegian context (at least at the central level). Nonetheless, despite attempts undertaken by policymakers in different contexts to prevent perverse effects from occurring, a growing body of research highlights that effects such as curriculum narrowing or educational triage cannot be avoided by lowering the

stakes (see for example Barbana et al., 2019; Feniger et al., 2016; Maroy & Pons, 2019; Thiel et al., 2017). This dissertation contributes to this body of research by highlighting how, also in the Norwegian context, PBA policies can generate a strong focus on tested content, which sometimes seems to go at the expense of other educational purposes and content.

This finding has important policy implications. Namely, there exists a need to recognize that the current system seems to generate pressure to prioritize cognitive competencies over non-cognitive competencies. In this regard, an important role is played by some local authorities and media outlets in raising pressure to prioritize tested content. With regards to the latter, this dissertation's findings highlight how not publishing school results on the 'School Portal' is not a sufficient measure to prevent the media from constructing performance rankings, and, as such, from contributing to performative pressure. As long as media outlets are able to request school performance data following the Freedom of Public Information Act, there is reason to suspect they will do so. That is, as highlighted by Chapter 3, media interest in standardized test performance does not seem to fade away with time, while having to issue a formal request to obtain school data does not seem to form a barrier to do so. In this light, exemption of test data from the Freedom of Public Information Act might be the only way to overcome the construction of performance rankings by media outlets.

Simultaneously, it is crucial to recognize that the PBA system itself constructs new definitions of education quality and success, thereby reframing norms of good educational practice. That is, in important ways, the PBA system promotes test performance as a prime indicator of educational success. This dissertation highlights how politicians across the political spectrum, as well as media outlets and some local politicians and authorities, increasingly rely on test scores to make judgements about education quality. As such, both equity and quality have been rearticulated in important ways to performance indicators based on national and international tests. As argued by Falabella (2014, p. 4), there lies a major risk here, of an 'ethical transformation, as performance and competition become overriding purposes, marginalizing critical thinking, social inclusion values, and democratic principles'. While most principals interviewed for this study seemed able to counter the pressure to prioritize academic achievement, the findings simultaneously highlight how a new 'generation' of principals is standing up in Norway, who increasingly measure themselves according to performance metrics, and aim at moving up in the performance hierarchy. This finding aligns to research in other contexts, where scholars have documented how some principals and teachers increasingly embody performance metrics (e.g. see Brass & Holloway, 2019; Courtney, 2016; Hefferman, 2017; Holloway & Brass, 2018;

Lewis & Holloway, 2017). As such, it appears that performance metrics can generate new professional identities and subjectivities.

Thus, rather than lowering the stakes to prevent perverse effects from occurring, there might exist a need to re-think the PBA system, and possibly to move towards what is referred to as an ‘intelligent’ accountability system, which recognizes the broad purposes of education, as well as the crucial responsibilities of other actors in securing education quality and equity, including governments, local communities and parents (Lingard, Sellar, & Lewis, 2017). In this regard, it remains crucial to recognize that only limited aspects of the education system can be measured (cf. Spillane, 2011) and used as a basis for accountability purposes. Working towards broader aims according to the societal mandate of schooling, such as in the Norwegian case, lies therefore within key actors’ professional responsibility and is strongly linked with work ethics and sense of moral obligations towards the students. It is therefore worthwhile reflecting on the question whether re-thinking the PBA system requires a move back towards greater trust in the profession.

References

- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu-net: Globalisation and Education Policy Mobility*. London: Routledge.
- Barbana, S., Dumay, X., & Dupriez, V. (2019). Local implementation of accountability instruments in the French-speaking community of Belgium. *European Educational Research Journal*, 19(2), 94–108.
- Baroutsis, A., & Lingard, B. (2017). Counting and comparing school performance: an analysis of media coverage of PISA in Australia, 2000–2014. *Journal of Education Policy*, 32(4), 432–449.
- Bjordal, I. (2016). *Markedsretting i en urban norsk skolekontekst. – Et sosialt rettferdighetsperspektiv*. PhD dissertation, Norwegian University of Science and Technology. Retrieved from: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2428050>
- Booher-Jennings, J. (2005). Below the Bubble: “Educational Triage” and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231–268.
- Brass, J., & Holloway, J. (2019). Re-professionalizing teaching: the new professionalism in the United States. *Critical Studies in Education*. doi: 10.1080/17508487.2019.1579743

- Courtney, S. J. (2014). Post-panopticism and school inspection in England. *British Journal of Sociology of Education*, 37(4), 623-642.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173-189.
- Espeland, W. N., & Sauder, M. (2012). The Dynamism of Indicators. In K. E. Davis, A. Fisher, B. Kingsbury & S. E. Merry (Eds.), *Governance by Indicators* (pp. 86-109). Oxford: Oxford University Press.
- Falabella, A. (2014). The performing school: The effects of market and accountability policies. *Education Policy Analysis Archives*, 22(70), 1–26.
- Feniger, Y., Israeli, M., & Yehuda, S. (2015). The power of numbers: the adoption and consequences of national low-stakes standardised tests in Israel. *Globalisation, Societies and Education*, 14(2), 183-202.
- Gunnulfson, A. E. (2017). School leaders' and teachers' work with national test results: Lost in translation? *Journal of Educational Change*, 18(4), 495-519.
- Haugen, C. R. (2019). Teachers' experiences of school choice from 'marginalised' and 'privileged' public schools in Oslo. *Journal of Education Policy*, 35(1), 68-94.
- Heffernan, A. (2017). The accountability generation: exploring an emerging leadership paradigm for beginning principals. *Discourse: Studies in the Cultural Politics of Education*, 39(4), 509-520.
- Holloway, J., & Brass, J. (2018). Making accountable teachers: the terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361-382.
- Jacob, B. A., and Levitt, S. D. (2003). *Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. Working paper No. 9413. National Bureau of Economic Research. <http://www.nber.org/papers/w9413>
- Jarke, J., & Breiter, A. (2020). *The Datafication of Education*. New York: Routledge.
- Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227-247.
- Kassim, H., & Le Galès, P. (2010). Exploring Governance in a Multi-Level Polity: A Policy Instruments Approach. *West European Politics*, 33(1), 1–21.
- Le Galès, P. (2010). Policy Instruments and Governance. In M. Bevir (Ed.), *The SAGE Handbook of Governance* (pp. 142–160). London: SAGE Publications.
- Lewis, S., & Holloway, J. (2018). Datafying the teaching 'profession': remaking the professional teachers in the image of data. *Cambridge journal of education*, 49(1), 35–51.

- Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. In G. Noblit (Ed.), *Oxford Research Encyclopaedia of Education* (pp. 1-28). New York: Oxford University Press.
- Maroy, C. (2015). Comparing Accountability Policy Tools and Rationales: Various Ways, Various Effects? In H. G. Kotthoff & L. Klerides (Eds.), *Governing Educational Spaces: Knowledge, Teaching, and Learning in Transition* (pp. 35-56). Rotterdam: Sense Publication.
- Maroy, C., & Pons, X. (2019). *Accountability Policies in Education. A comparative and multilevel analysis in France and Quebec*. Cham: Springer.
- Mausethagen, S. (2013). Talking about the test. Boundary work in primary school teachers' interactions around national testing of student performance. *Teaching and Teacher Education, 36*, 132-142.
- Mausethagen, S., Prøitz, T. S., & Skedsmo, G. (2017): Teachers' use of knowledge sources in 'result meetings': thin data and thick data use. *Teachers and Teaching, 24*(1), 37-49.
- Prøitz, T. S., Mausethagen, S., & Skedsmo, G. (2019). District administrators' governing styles in the enactment of data-use practices. *International Journal of Leadership in Education, 1-22*.
- Reback, R. L. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics, 92*(5-6), 1394-1415.
- Sahlberg, P. (2016). The Global Education Reform Movements and its Impact on Schooling. In K. Mundy, A. Green, B. Lingard & A. Verger (Eds.), *The Handbook of Global Education Policy* (pp. 128–144). Oxford: Wiley Blackwell.
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). Evaluering av nasjonale prøver som system [Evaluation of national testing as a system]. Published online: <https://www.nifu.no/publications/1025043/> Accessed 1 February 2020.
- Skedsmo, G. (2018). Comparison and Benchmarking as Governing Processes in Norwegian Schools. In L. M. Carvalho, L. Lévassieur, M. Liu, R. Normand, & D. A. Oliveira (Eds.), *Education Policies and the Restructuring of the Educational Profession* (pp. 137-158). Singapore: Springer.
- Spillane, J. P. (2012). Data in Practice: Conceptualizing the Data-Based Decision-Making Phenomena. *American Journal of Education, 118*(2), 113–141.
- Statistics Norway. (2019). Population and Land Area in Urban Settlements. Statistics Norway. <https://www.ssb.no/en/befolkning/statistikker/befteft>
- Steiner-Khamsi, G. (2016). New directions in policy borrowing research. *Asia Pacific Education Review, 17*, 381–390

- Theodorsen, G. (2017). Hvordan opplever erfarne lærere på 1.-4. trinn at skolen deres har endret seg de siste 15 årene? Master's degree dissertation, University of Tromsø.
- Thiel, C., Schweizer, S., & Bellmann, J. (2017). Rethinking side effects of accountability in education: Insights from a multiple methods study in four German school systems. *Education Policy Analysis Archives*, 25(93), 1–32.
- Verger, A., & Parcerisa, L. (2017). *Accountability and education in the post-2015 period: International trends, enactment dynamics and socio-educational effects*. Think piece for the UNESCO GMER 2017. Paris: UNESCO.
- Verger, A. Parcerisa, L. & Fontdevila, C. (2019). The Growth and Spread of National Assessments and Test-based Accountabilities: A Political Sociology of Global Education Reforms. *Educational Review*, 71(1), 5–30.