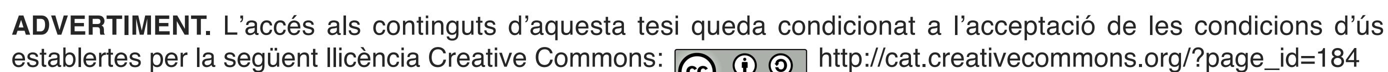
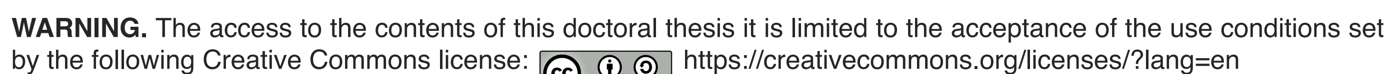
Universitat Autònoma de Barcelona


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

HOLDING SCHOOLS ACCOUNTABLE

DEVELOPING & ENACTING TEST-BASED ACCOUNTABILITY IN THE AUTONOMOUS DUTCH EDUCATION SYSTEM

NATALIE BROWES

Universitat Autònoma
de Barcelona

·PhD Programme in Sociology·

·Department of Sociology·

·Faculty of Political Science and Sociology·

Holding Schools Accountable: Developing and Enacting Test-Based Accountability in the Autonomous Dutch Education System

~ Doctoral Dissertation by ~
Natalie Browes

Directed by

Dr. Antoni Verger Planells

Dr. Hülya Kosar-Altinyelken

July 2021

This research was supported by the H2020 European Research Council [StG-2015-680172].

~Acknowledgements~

This thesis could not have been completed without the help and support of a considerable number of individuals and organisations. For this, I am very grateful.

I would like to firstly express my gratitude to all of the participants involved in this research for their interest and openness. Despite their busy schedules, over sixty individuals from the fields of education policy and practice generously shared their time, experience and insights, and always in good spirits. From them, I learnt a lot. I am also grateful to those who supported data collection and helped me access participants, and in particular, to Wiljan Hendrikx, for generously sharing his knowledge, enthusiasm, and contacts.

I would like to say a big thank-you my supervisor, Antoni Verger, for his impressive expertise and invaluable guidance over the last four years and for developing such an innovative and ambitions project. Equally, my thanks go out to Hülya Kosar-Altinyelken, my supervisor in the Netherlands, for her continuous encouragement, vital insights into the Dutch system, and for once again taking on the task of overseeing my work. Toni and Hülya, you have always made yourselves available, always shown kindness, and have motivated and supported me throughout the long process. My colleagues at both the Autonomous University of Barcelona and the University of Amsterdam have taken on a similarly motivating role, being generous in the time they have dedicated to reviewing early drafts of various chapters. My ReformEd colleagues; Andreu, Anto, Clara, Edgar, Gerard, Guilia, Isabel, Laura, Lluís, Marcel, Marjolein and Mauro, always impress me with their sharp insights and impressive subject knowledge, and developing the project together with them has been extremely rewarding. I have shared my PhD journey particularly closely with Marjolein, and I thank you for all your support and generosity over the years. Likewise, my UvA colleagues; Hester, Lina, Mariella and Semiha, have provided me with valuable and diverse perspectives from different aspects of the field of education, undoubtedly improving my own work, for which I am much appreciative.

My gratitude also extends to the two universities who have supported me throughout my PhD: the UAB for welcoming me, and particularly to those individuals, most recently Oscar Ruíz, who have generously helped me navigate various bureaucratic challenges; and the UvA for being a generous partner and hosting me during my time collecting data and beyond. Further, thank-you to the European Research Council who made this research and the wider project possible through its funding.

Finally, thank-you to all my family and friends who have been there over the past years. In particular: to James – thank-you for supporting me in so many ways, enabling me the freedom to follow my interests; to my sister – thank-you for your continuous encouragement and for taking the time to listen; and, to my parents – thank-you for the proofreading, the childcare, and for all the other ways you have supported me that there is not the space here to mention.

Dedicated to Jack

TABLE OF CONTENTS

CHAPTER 1

Introducing the Research ... 1

Thesis Overview ... 1

Theoretical Framework ... 3

A Global Policy in a National Context ... 3

Teachers & TBA ... 5

Contextual Overview: Why the Netherlands? ... 7

Relevance & Rationale ... 8

Methodological Framework ... 9

Table 1: Summary of Research & Participants ... 12

Thesis Structure ... 13

Compendium of Publications ... 14

References ... 15

CHAPTER 2

A Global Policy in a National context – *The instrumentation of test-based accountability in the autonomous Dutch system* ... **19**

Introduction ... 19

Adopting a Systemic Policy Instruments Approach to TBA ... 21

Table 2: GERM Principles & the Role of National Assessments ... 22

Methods ... 24

National Context ... 26

Findings: The Development & Evolution of TBA in the Netherlands ... 28

The Development of TBA ... 28

Conditions for Reform & the Reception of Global Ideas ... 31

The Translation of TBA in an Autonomous System ... 34

Discussion ... 36

Conclusions ... 38

References ... 39

CHAPTER 3

SAWA & GLOBAL GOVERNANCE - *The Dutch Experience of OECD Reviews of National Policies for Education* ... **44**

Introduction ... 44

The Ideas Game: The Role of Policy Reviews in the Transnational Governance of Education ... 46

Policy Reviews as a Form and Practice of Institutional Power ... 46

The Politics of OECD Policy Reviews ... 47

Policy context: The OECD Policy Agenda and Dutch Education 49

The OECD Agenda for Education 49

Education Policy in the Netherlands: The 'Good Student' 50

Methods 52

Findings 52

Reasons for Commissioning the Review 53

Establishing and Negotiating the Agenda and Review Process 54

'Problems' and 'Solutions': A focus on SAWA policies 56

Use and Impact of the Review 58

Discussion and Conclusions 60

References 62

CHAPTER 4

Policy Enactment I - *Professionalism in the Era of Accountability: Role Discrepancy & Working Strategies amongst Teachers in the Netherlands* **66**

Introduction 66

Professionalism, Role Discrepancy & Enactment 68

Teachers and Accountability in the Netherlands 70

The Dutch Teaching Profession 70

Accountability & Performativity 71

Methods 72

Table 3: Participating Teachers & their Schools 74

Findings 74

Teachers' Tasks: Practices and Beliefs 74

Role Discrepancy & Teacher Response 77

A Shifting Sense of Professionalism? 80

Discussion and Concluding Thoughts 83

References 85

CHAPTER 5

Policy Enactment II – *Test-based Accountability & Perceived Pressure in an Autonomous Education System: Does School Performance Affect Teacher Experience?* **89**

Introduction 89

TBA, Performativity & Enactment Theory 91

State of the Art 93

Accountability in Dutch Schools 94

Methods 96

Table 4: List of Participants 98

Findings 98

Performance Context & Logics of Action: A Study of Four Schools 99

Teachers' Experiences of Performativity & Pressure 103

Discussion 109

Conclusions 111

References 112

CHAPTER 6

Global Summary of Findings 116

The Evolution of School Accountability in the Netherlands & the Role of School Autonomy 116

The Role of International Organisations in the Promotion of SAWA &TBA 117

The Impact of TBA & Performative Work Environments on Teachers' Practices, Beliefs & Professional Identities 118

Teacher Experience of TBA & Performance Pressure in Low & High Performing School Contexts 120

References 122

CHAPTER 7

Conclusions 124

Discussion of Main Findings 124

Limitations & Further Research 126

Recommendations for Policy & Practice 127

Final Thoughts 130

References 130

APPENDIX

Table 6: Data collection & Participant Characteristics 132

CHAPTER 1

INTRODUCING THE RESEARCH

THESIS OVERVIEW

Despite being a country associated with contented citizens (Helliwell, Layard & Sachs 2019), an enviable work-life balance (Eurostat 2019), and a stable, consensus-building political approach (the famous 'polder model'), over recent years, one sector of Dutch society has been experiencing considerable unrest, repeatedly described as being in a state of 'crisis'. Since the summer of 2017, the education sector has seen recurrent strike action. Started by primary teachers, supported by teacher unions, and more recently joined by their secondary school counterparts, this, previously rare, occurrence has resulted in school closures across the country. Teachers' widespread frustrations have come to a head, largely driven by the movement 'PO in Actie' – demanding a seat at the policy table for primary teachers and a change in working conditions. In particular, these frustrations comprise heavy workloads, increasing levels of stress, and insufficient pay. To a large extent, work pressures are attributed to, yet also reflected by, a significant teacher shortage. This issue is particularly prevalent in large urban areas: the profession, it seems, is no longer attracting and retaining the workforce it requires. Media reports detailing classes being sent home and entire school closures (either temporarily[1] or permanently[2]) are becoming commonplace. But beyond needing a few extra staff and a little extra money, these trends reflect deep seated, structural shortcomings that call for meaningful systemic change. As two Dutch teachers put it: what we need is to 'flip the system,' to change education from the ground up, with teachers positioned firmly at the centre of the process (Evers & Kneyber 2015).

This thesis looks behind the curtain of this unrest, to try and understand the causes of some of these issues and explore how they are experienced by teachers. It examines a set of global education policies through their introduction, evolution and enactment in the context of the Netherlands. These are school

[1] See https://www.trouw.nl/nieuws/16-amsterdamse-scholen-week-dicht-om-lerarentekort~bf2f4a95/. Last accessed on 16th October 2020.

[2] For example, see https://nos.nl/artikel/2303487-basisschool-amsterdam-dicht-om-tekort-aan-leraren.html. Last accessed on 16th October 2020.

autonomy with accountability, or 'SAWA' policies and are at the heart of what has ominously become known as the 'GERM' (the Global Education Reform Movement) (see Sahlberg 2016). As a structure of educational governance, SAWA aims to increase efficiency and effectiveness by making schools and/or their management bodies responsible for school quality (autonomy), while keeping system quality 'in check' through the introduction of attainment levels, evaluated through standardised assessments to which various consequences may be attached (accountability). This policy model is particularly promoted by influential international organisations (IOs), including the Organisation for Economic Cooperation and Development 'OECD' (see for example OECD 2011), and the World Bank,[3] primarily as a way to raise student performance. The thesis is the product of four years of doctoral research as part of the 'ReformEd' research team based at the Autonomous University of Barcelona, investigating these policies in different educational settings with diverse institutional traditions, including; Chile, Spain, Brazil, Italy and Norway. In the Netherlands, the research has been supported by the University of Amsterdam.

The research focuses in particular on the accountability – known as test-based accountability, or 'TBA' – aspect of this policy package, and its impacts. This is an area that has received considerable academic attention, particularly in Anglo-Saxon early adopters; the US, England, and Australia, yet, there is still a lot to be discovered. While TBA policies may look very similar from one country to the next, they in fact differ in important ways. This is particularly true in education systems such as the Dutch one, which are based on quite different institutional traditions to their Anglo-Saxon counterparts. Insufficient research into these systems means that it is still unclear why such policies are introduced, how they evolve, and furthermore, which formulations of accountability produce which effects.

To investigate this, the research adopts a broad yet in-depth perspective to understand SAWA. It recognises the various drivers and influencers of policy, and traces the policy process from adoption, to translation, to enactment. On the one hand, it is understood that national policy has become a global business, fuelled by IOs through data gathering, information sharing, and competition. On the other hand, it also appreciates that education policies are shaped by a unique set of cultural, political and economic factors at the (sub) national level. Understanding such policies therefore requires an examination of both global and local dynamics, and the roles of local, national and transnational actors. Currently, this multi-level approach is lacking from the field. Research in the past has predominately focused either at the macro level – studies informed by World Culture Theory that neglect the enactment of policy, or at the micro level – ethnographic studies that do not make empirical links at the national and supranational levels (Verger 2014). By focusing on contextual particulars whilst recognising global

[3] See for example: http://saber.worldbank.org/index.cfm?indx=8&pd=4&sub=0 (last accessed 30th September 2020)

mechanisms and using research tools that enable comparisons and generalisations to be made, the thesis and its wider project adopts what is known as a systems theory approach (Luhmann 1990).

Research for the study has been divided into two strands. The first, aims to understand policy adoption and translation at the national level. The second, focuses on the meso and micro levels, exploring policy enactment and the impact on teachers and the teaching profession more broadly. While the strength of the wider project lies in its development of (contextualised) comparative research tools, and a common methodological approach, it is important to note that, in order to explore and analyse the case effectively, this thesis focuses exclusively on the Netherlands. It seeks to answer the research question below, subsequently broken-up into sub-questions, each of which is addressed by one chapter of the thesis:

How and why have school autonomy and test-based accountability policies evolved in the Netherlands over recent years, and what impact has this had on teachers' practices, beliefs and overall professional experiences?

1) *How has school accountability policy evolved in the Netherlands, and what role has school autonomy played within this?*
2) *What is the role of international organisations in the promotion of SAWA and TBA?*
3) *In performative environments shaped by TBA, how are teachers' practices, beliefs and professional identities impacted?*
4) *What is the effect of TBA on differently performing schools and (how) does this affect teachers' experiences of their work and perceptions of performance pressure?*

THEORETICAL FRAMEWORK

A Global Policy in a National Context

Research strand one seeks to uncover multi-level mechanisms that have shaped TBA in the Netherlands. It applies theories that have sought to explain why we see strong similarities in education policy across the globe, how these policies are shaped to the local context, and how policy can be influenced by global actors that have no direct legislative power in the nation state. While TBA in the Netherlands is further explained within the introductory and contextual sections of subsequent chapters, for now it is perhaps relevant to provide a brief overview of these policies, how they are designed to operate, and the assumptions on which they are based.

The SAWA model of educational governance is guided by new public management and private sector values of efficiency and effectiveness. Through a model of decentralisation, standardisation and administrative and market forms of accountability, governments are able to take a less direct, more economically efficient approach to ensuring system quality in education. Standardised testing lies at the centre of this model, bringing together these core principles (see table 2 chapter 2). The tests themselves

tend to measure centrally defined standards, usually in the 'core' learning areas – numeracy and language. Becoming a proxy of education quality, (published) test results serve to stimulate competition and school improvement. SAWA, it is argued, drives up standards and narrows educational inequalities: offering 'objective' measures of student performance and better allowing schools to adapt teaching to the local context and to become more responsive and innovative (see Coleman et al 1997). The model is expected to result in greater community engagement and increase feelings of ownership over education provision, thereby strengthening relationships between teachers, and students and parents, and increasing job satisfaction (see for example, OECD 2013; World Bank 2016).

Increased autonomy through decentralisation became commonplace in education, as in other sectors, in the neoliberal era of the 1980's and 90's. Led by Thatcher in England and Reagan in the US, this period saw a shift from government to governance. Greater autonomy over the hiring of staff, spending of finances, and pedagogical and curricular decisions was promoted to enable schools the room to diversify and better respond to the various needs of their student populations. This was often accompanied by privatisation, school choice, and funding that followed the pupil – an early form of market accountability (see for example, the UK parliament's Education Reform Act of 1988). Accountability tools expanded and became refined over the years, staying with the example of the UK, the education act of 1992 saw the publishing of school results in England. In the US accountability hit hard under Bush Junior with the introduction of the 'No Child Left Behind Act' in 2002. Operating at various levels (state, district, school, and teacher) administrative tools introduced a range of rewards and sanctions that were attached to student test performance, from salary bonuses to school closures. Many of these accountability measures are based on the fundamental assumption that teachers and schools play an essential role in student outcomes (Tatto 2007).

This model of education governance has now spread well beyond its Anglo-Saxon roots. While varying in important details, broad similarities can be seen in the policy instruments being adopted within vastly different education systems across the globe. The work of policy borrowing theorists has been particularly useful in understanding the transference of policy across education systems that results in national variances (see the work of Steiner-Khamsi). One important reason for this global spread is, from a government perspective, the relative ease of implementation: rather attractively, SAWA places the burden of quality maintenance away from governments and onto schools and teachers. Another, is the flexibility of the model and its component parts, described as an 'empty vessel' (Steiner-Khamsi 2014; 2016) that can be employed by the political left or right to address a vast array of problems (Verger, Parcerisa & Fontdevila 2019).

Beyond SAWA characteristics, policy borrowing also calls for a contextualised reading of the nature of, and reasons behind, policy adoption – a process known as 'reception' (Steiner-Khamsi 2014; 2016). Here, the influence of transnational stakeholders and their role in global education governance must also

be acknowledged. In particular, much has been written about the OECD and the vast reach and impact of its Programme for International Student Assessment (PISA) (see Breakspear 2012; Grek 2009). Yet other modes of influence, including the organisation's use of policy peer reviews, are under-researched. Peer reviews can be understood as practices revolving around 'multilateral surveillance' (Porter & Webb 2008; Schuller 2005): subtle power constituting indirect forms of coercion that explain how SAWA policies are promoted and why a national government might engage with this (see chapter 3).

Moving beyond the *reception* of policy, it is equally pertinent to analyse the process of policy *translation* (Steiner-Khamsi 2014), to discover: which parts of these global policies have been 'borrowed', how have they been shaped or recalibrated to the national context, and why. A policy instruments approach is particularly useful here, interpreting policy through the unique set of tools which constitute it (see Le Galès 2010). It holds that such tools are not simply selected for their effectiveness, rather, also looks for the cultural, political and economic reasons behind tool choice (see chapter 2).

Teachers & TBA

Several studies have dedicated themselves to testing the assumptions on which the SAWA model is based, and revealing the impacts of these policies. Quantitative studies have tested the underlying claim that SAWA results in higher student achievement, yet, results have proven mixed or inconclusive (Hanushek, Link & Woessmann 2013; Ko, Cheng & Lee 2016; Woessmann, Lüdemann, Schütz & West 2007). Studies focusing on the link between student achievement and high-stakes testing have similarly proven inconclusive (Hanushek & Raymond 2004; Nichols Glass & Berliner 2006). Taking a less instrumental approach to education, another field of research has sought to understand the enactment and impact of these policies on schools, students, and teachers. It is these findings, particularly those at the level of the teacher, which are most pertinent to this thesis.

Enactment theory helps shine a light on how TBA policies are interpreted and translated by practitioners. The influential empirical research that shapes these theories has largely been conducted in the US, UK, and Australia (see the works of academics such as James Spillane, John Diamond, Heinrich Mintrop, Stephen Ball, Meg Maguire, Annette Braun, Bob Lingard …), yet research on the enactment of TBA is now expanding into different education systems. Complementing the broader goals of this study and its wider theoretical and methodological approaches, enactment theory gives prominence to context. Studies have highlighted the importance of policy context (the stakes attached to testing) (Nathaniel, Pendergast, Segool, Saeki, & Ryan 2016), market context (Maroy & Van Zanten 2009) and school context – particularly school population and performance level (Diamond & Spillane 2004). Findings have revealed a host of undesirable effects on teachers and theories have pointed to a teaching profession that is fundamentally changing.

Findings, in particular those from the US and the UK, have revealed that narrow quality standards, prescribed curricula, and high stakes accountability (linking results to 'rewards' such teacher bonuses

and 'sanctions' such as teacher probation, dismissal, and even school closure) and have resulted in a host of undesirable effects. These are not only changing teaching practices, but changing the very notion of what it means to be a good teacher. Teachers, it is claimed, now work in environments of 'performativity' (Ball 2003) and are required to become 'technicians' (Taylor 2007). Such research has shown time and again that teachers feel constricted, pressured and overworked (for example, Valli & Buese 2007). With accountability pressures affecting teacher autonomy – dictating what and how they teach – and a perception that this negatively impacts their students' experience (Berryhill, Linney & Fromewick 2009; Valli & Buese 2007), it has also been found that teachers are left torn between policy demands and professional values (Braun & Maguire 2018; Hendrikx 2019). Despite this, research into individual's experience of 'role discrepancy' - the incongruence between their [the professional's] ideal roles and the roles they actually engage in at work' (Takase, Maude & Manias 2006 p.752) is lacking in the teaching profession (see chapter 4). Further, studies have also indicated that undesirable side effects are particularly prevalent in low-performing and probation schools, often attended by high percentages of students for whom English is not a first language. Yet these studies have tended to exclusively focus on low performing schools (Mintrop 2003), or have structured comparisons between low and high performing schools around the premise of 'school response' (Diamond & Spillane 2004) and 'school improvement' (Mintrop 2007). The studies have also limited their focus to administrative accountability. There remains a gap in understanding about the various sources of pressures operating on schools, how teachers' work and identities are impacted by school-level responses, and whether better school performance results in reduced teacher pressure.

More recent research, including pieces from this project (Camphuijsen 2020; Pagès 2021), shows that such undesirable side effects at the school and teacher level may not be exclusive to high-stakes environments. Thiel & Bellmann (2017) for example, also found adaptive behaviour (the assimilation of teaching) in German states classed as adopting low or no-stakes accountability, while Penninckx, Vanhoof, Maeyer, & Petegem (2016), investigating the effects of inspection in the low-stakes context of Flanders, found strategic behaviour previously associated with high-stakes inspection regimes, to be common within schools.

In such ways, accountability demands and performance pressures are reshaping teachers' work and their very identities (Day 2002; Hargreaves 2000). Teachers can be said to be experiencing a process of 're-professionalisation'. Engaging with literature on the sociology of the professions, these changes are seen to be representative of broader, macro-level shifts that are taking place: from professionalism defined by collegiality, trust and autonomy, to professionalism defined by managerialism, standardisation, and external accountability (Evetts 2009). Not only is this the result of ongoing neoliberal reforms, but it is the result of the intrusion of neoliberalism into the core of society and the psyches of individuals (Ball 2016). How teachers feel about and respond to this is an important question to ask. Research has indicated that teachers may be adapting to these new professional realities, with those who are newer to

the profession seemingly (better) adjusting themselves to TBA principles (Holloway & Brass 2018; Wilkins 2011).

CONTEXTUAL OVERVIEW: WHY THE NETHERLANDS?

Given the prevalence of SAWA in countries across the globe, it is perhaps pertinent to explain why the Netherlands was chosen as a case study to further examine this policy phenomenon. The Netherlands is one of many cases where significant questions remain about the evolution of accountability, its influences, and its impacts. Perhaps due to the well-reputed nature of education (performing well overall in international indicators) (OECD 2017), and its highly-decentralised system with government control over standards and testing, the Dutch 'version' of SAWA has already received attention internationally (OECD 2016; UNESCO & GMR 2017; World Bank 2011). Yet these discussions remain largely normative, and lack a broader, critical perspective on policy development and effects.

The Dutch system is intriguing for a number of reasons. In many ways, it reflects the 'global' SAWA model: increased decentralisation taking off in the 1980's/90's and continuing through to the mid noughties, gave boards and their schools increased responsibilities in key areas such as budget management (see Karsten 1999). This was followed by an intense period of accountability-based reform (ca.2008-2014), which saw the introduction of core learning standards, compulsory standardised testing, and stakes attached to test performance. Yet at the same time, the education system we see today is the product of a unique institutional legacy. As further explained in chapter 2, school autonomy did not originate in the neoliberal reforms of the 80's and 90's, rather, has been built into the system for over 100 years. As a result, the Netherlands is classed as one of the most autonomous systems in the world (Nusche, Braun, Gábor, & Santiago 2014), and school(board) freedoms are fiercely protected. 'Freedom of Education' has resulted in a highly diverse system, and is still pivotal in shaping policy today. The impacts that TBA might have in such a (theoretically) autonomous and diverse system is a question that holds great personal interest, and one that has driven this research over the past four years.

Within the Dutch system, a decision was made to focus on primary education (ISCED 1). As remarked by Braun and Maguire (2018), primary school dynamics differ in important ways to those in secondary education, tending, for example, to be smaller and the relationships with staff and management often more personal. Furthermore, primary schooling is more female dominated, and teachers spend considerably more time with their students and form closer relationships. This, it is noted, 'can cast primary teachers in a 'maternal' caring role, and they may struggle to assert professional autonomy' (Braun & Maguire 2018 p.4). The decision to focus on primary education was made despite the fact that the wider ReformEd project focuses on both ISCED levels 1 and 2. This is because standardised testing does not have the same function in Dutch lower secondary education, rather, accountability-

based reform has been particularly prevalent at the primary level. Despite this, limited research has focused on TBA and teachers in the Netherlands, and even less within primary education.

RELEVANCE AND RATIONALE

Despite the prevalence of TBA across education systems and the wealth of studies that have examined some of the effects of these policies, there is still much that is not known, or that remains under-researched. This thesis aims to address some of these research gaps, and in doing so, make a small academic contribution to the field. Further, given the very real-life issues addressed by the research, as indicated in the opening paragraph of the thesis, it also aims to make a societal contribution, going some way to help inform future education policy and practice.

- The research takes a multilevel approach to understand policy development and impact, combining macro theories on policy transfer with empirical case-based data. By critically examining the evolution of school autonomy and (test-based) accountability in one system, it seeks to contribute to academic knowledge concerning how and why 'global policy' is adopted and adapted, and fundamentally, how its evolution is transformed by national context and institutional history.

- One way in which this global-national relationship is more closely examined, is through the role of OECD peer reviews. This is an under-researched area. Studies that have taken place have been in systems that are contextually quite different to the Netherlands, and are generally limited to the more 'extreme' cases where reviews have had considerable impacts. This research add value by uncovering new factors that explain engagement with, and impact of, peer review in the case of one 'good student.'

- One considerable research gap in the TBA literature, highlighted within the theoretical framework section, is that left by the dominance of research in traditionally high-stakes contexts, particularly the neoliberal, Anglo-Saxon administrative models described. More data is needed from contexts where administrative traditions and formulations of TBA differ. From the limited research that has been conducted in 'lower-stakes' systems, it is suggested that undesirable side effects (school gaming strategies and teacher adaptive practices) may persist. Yet more can be done to explore this, particularly in the form of qualitative research that examines not only objective practices, but teachers' feelings about their work, their perceptions of pressure, and their professional identities. By focusing on these issues in the unique Dutch context, this research adds to our understanding of which formulations of TBA produce which effects.

- The research places further emphasis on context by also gathering data in different school settings within the same policy context. As noted, whereas research has identified the particular challenges facing teachers in low-performing schools that, by nature of the accountability system, find themselves under greater administrative demands (for example, Mintrop 2003), fewer studies directly compare these challenges with those faced by their counterparts in high-performing schools. Those that do, focus on school responses to TBA rather than teacher experiences of it (Diamond & Spillane 2004). Therefore, by conducting comparative research in these different settings, the thesis attempts to provide a deeper understanding of the circumstances surrounding particular TBA effects, asking: 'what works, where, and for whom?'

- Beyond exploring new contexts, the thesis makes a conceptual contribution in its application of the notion of role discrepancy to the teaching profession and its identification of various responses to this. By doing so it complements but also builds on work that has been conducted in the Netherlands at the secondary school level (see Hendrikx 2019), and moreover, is able to make inter-professional comparisons and draw conclusions that extend beyond the field of education.

- The deeper, more nuanced understandings of policy development and enactment that this thesis strives towards are crucial to better understand the experiences, responses, and potential struggles of teachers. In the Netherlands, a teacher shortage is a significant and ongoing issue, and incidences of teacher burnout are higher than in other sectors of the Dutch labour market (Inspectorate of Education 2019). These issues have been found to be particularly prevalent in lower-performing schools, which risks exacerbating student inequalities. Recent years have seen repeated strike action from primary school teachers, fuelled by high workloads, significant work pressures, and perceptions of low and unfair pay. A better understanding how policy reform has affected teachers' daily practices and how they feel about their work is central to help relieve some of these pressures on teaching professionals and the education system more broadly.

METHODOLOGICAL FRAMEWORK

Given the focus that the research places on context and the importance of (hidden) mechanisms, a realist evaluation approach has been adopted to guide data collection and analysis (see Pawson & Tilley 1997). This approach highlights questions such as: *how and why has policy developed as it has?* And: *What impact does this have, to what extent, on which actors, in which circumstances and why?* To answer

such questions, the Netherlands is taken as a case study (Yin 2009). Two points are important here: firstly, while this study is focused on one national case, it is not bound by national boarders, rather, as clarified throughout the thesis, giving attention to the role of transnational trends and dynamics is crucial to understand education policy in the Netherlands. Secondly, while the 'case' in question is defined as the nation, this does not mean that findings at the policy-enactment level are representative of the country. Rather, the qualitative research is designed to provide an in-depth insight into the impact of test-based accountability on schools and teachers in a small number of different settings. While work on more representative data is ongoing in the form of a nationwide survey, as further explained in the 'limitations and further research' section of chapter 7, this complementary data does not form part of the thesis.

The data underpinning this study was collected between October 2017 and November 2019. The first two research questions, which focus on the policy-making level, are answered through two sources of data; documents analysis and expert interviews. The final two research questions, concerned with policy enactment, are tackled primarily (but not exclusively) through interviews with practitioners. Question 3 relies on in-depth interviews with teachers from six schools, while question 4 is answered through a focus on four of these schools that fall under the same management, also including interviews with the school principals, the head of the school board, and analysis of key school documents. For an overview of data collection, see table 1.

'Experts' were purposefully selected based on their role (Patton 2015), and identified through background research. Subsequent to this, snowballing techniques were used (Teddlie and Yu 2007). Regarding school selection, public and independent publicly-funded primary schools qualified for the study. Privately-funded schools did not, given that they are not subject to the same government regulations on testing and accountability. Schools were selected that are located in small cities in four neighbouring provinces in the Randstad, a highly urban region in central-west Netherlands. Urban areas – where education markets are competitive – were chosen so that the impact of market accountability could be observed, yet the main Dutch cities were avoided given that a preliminary scanning of the field indicated over-research in these areas, and teacher shortages and high workloads to be particularly severe. Research question 4, calling for comparative data at the school level, required additional selection criteria. This included school performance (based primarily on published average student test scores) and school management (all schools under the same board to limit variability). Finally, teacher selection is weighted towards those in the upper years where testing plays a more significant role. Depending on the preferences of school principals, some teachers were contacted directly by the researcher while others were approached by the principal (once selection criteria had been explained). These teachers would be encouraged, but ultimately would elect whether or not to participate. A degree of self-selection bias was therefore unavoidable. For an overview of research participants, see table 1

at the end of this section. For a full list of participants and key characteristics, see table 6 in the appendix section.

Interview scripts and coding protocols were developed with the research team as part of the broader project, and contextualised and adapted individually according to the specific focus of this study. Interviews were semi-structured, enabling room for flexibility to incorporate participant experience and response. Expert interviews were longer than interviews with practitioners, averaging around 1 hour, and 40 minutes respectively. For both research strands, interview and coding protocols were primarily led by the study's research questions. While initial coding can thus be described as adopting a structural approach (Saldaña 2009), the process was iterative, with subsequent rounds focusing on the development of analytic codes (Deterding & Waters 2018) that reflected the particulars of the Dutch case. Coding was initially conducted in parallel by different members of the research group, enabling codes to be clarified or adjusted if needed, therefore limiting coder subjectivity. Background documents served different purposes according to the research strand and were thus analysed in different ways. Depending on their purpose, they were either coded using the same protocols as the corresponding interviews, or were treated in a more unstructured way and read for background information and to corroborate interview data. More information on this is available in the methods sections of subsequent chapters.

This study has been approved by the ethical committees of both the Autonomous University of Barcelona – the university hosting the wider project, and the University of Amsterdam – the partner university supporting the research in the Netherlands. The project was explained to all participants, their consent obtained in writing, and the possibility of their withdrawal at any point was made clear to them. Experts in the study, given their niche positions, have been described broadly by role only, and all schools and teachers have been given pseudonyms. Any other details that may reveal participant identity have been omitted.

As a researcher, it is also important to reflect on my own position and background, and how this inevitably affects my collection and reading of the data. Primarily, it is important to note that I did not grow up in the Netherlands and therefore am an 'outsider' to its education system. This was found to have an impact: particularly during expert interviews, being seen as an outsider can in many ways be beneficial. Participants were open and generous in their explanations of (their perceptions of) Dutch education and education policy, and did not seem concerned that I had any predetermined opinion of these matters, nor that I was associated with any particular Dutch organisations. The international nature of the project was also an aspect that was found to peak interest and encourage engagement, with participants attracted by the comparability function of the broader study.

Yet, this outsider position also comes with potential challenges. In particular, I want to briefly discuss the language in which data collection took place. Although words, phrases, or explanations were often

given in Dutch, the bulk of interviews were conducted in English. While all participants were offered the option of fully Dutch interviews (with the accompaniment of a native Dutch researcher), only one requested this. The level of English proficiency amongst participants was high, yet still it must be recognised that explanations may have been more precise, or conversations a little freer if conducted in their first language, particularly regarding more abstract, introspective questions. Documents analysed were read in both their original Dutch and their English translated form, and native Dutch speakers were consulted throughout the process.

As well as impacting participant response, my background also affects my own response. Conscious efforts were made to avoid shaping the data around any expectations. While research is inevitably based on underlying assumptions, these assumptions were tested throughout the data collection and analysis process; probing and follow-up questions were regularly used in interviews to clarify the meaning of responses (Ritchie, Lewis, Nicholls & Ormston 2013), school and teacher data was anonymised prior to transcription to minimise unintentional bias during the data analysis process, and, as mentioned, interview scripts and coding protocols were all developed in consultation with fellow researchers, helping to prevent an overemphasis of particular areas or a neglect of others.

Table 1: Summary of Research & Participants*

Research Strand	Question Addressed	Data collection	Participants
1	How has school accountability policy evolved in the Netherlands, and what role has school autonomy played within this?	25 interviews with Policy experts Analysis of policy documents	31 participants: policy-makers, policy consultants, policy-designers, policy advisers, academics, and representatives of key national sector organisations (council of school boards, teacher unions, teacher, parent and student organisations, national testing organisation, national curriculum development organisation)
	What is the role of international organisations in the promotion of SAWA and TBA?	Interviews with members of OECD policy review team and Dutch MoE Document Analysis	5 participants: 2 members of MoE and 3 members of OECD review team
2	In performative environments shaped by test-based accountability, how are teachers' practices, beliefs and professional identities impacted?	20 interviews with teachers across 6 schools	20 participants: 20 teachers
	What is the effect of TBA on differently performing schools and (how) does this affect teachers' experiences of their work and perceptions of performance pressure?	20 interviews in 4 schools* and with school board Analysis of school documents	20 participants: 1 head of school board, 4 (vice)principals, 15 teachers
TOTAL			**61**

* For a more detailed overview of research methods and participant characteristics, see table 6 in the appendix section.

THESIS STRUCTURE

Following this *CHAPTER 1* introduction, each core chapter of the thesis addresses one of the study's research questions. A compendium of publications is also included at the end of this chapter. Chapters two and three address questions at the level of policy-making, seeking to understand how and why school autonomy and test-based accountability evolved as it did in the Netherlands. *CHAPTER 2* approaches this query comprehensively, examining the development of a global policy in a 'highly-autonomous' national context. It recognises the importance of international ideas but also the unique national circumstances that shaped the timing of these policies and the form they took. *CHAPTER 3* enquires into a much more specific domain of policy influence, asking what role IOs play in the spread of SAWA. To do this, it focuses on the OECD and the organisation's use of policy peer reviews, following one such review recently conducted in the Netherlands. Drawing on a number of expert interviews with individuals from the OECD and the Dutch Ministry of Education (MoE) who were heavily involved in the review, it reveals the mechanisms through which policy is influenced, and how such reviews may be used by the OECD and also by national government to advance their various policy agendas. Chapters four and five move further down the policy line in an attempt to understand the enactment of TBA. *CHAPTER 4* explores the impact of TBA and the performative cultures and administrative burdens that accompany it, on the practices and professional identities of teachers. It asks whether teachers, as found in other professions, experience role discrepancy and if so, what is their response? *CHAPTER 5* examines four school contexts to better understand policy impact on teachers. Located in the same municipality and managed by the same school board, these schools vary considerably in their performance levels. As well as teachers, interviews were conducted with the school principals and with the director of the school board in recognition of the accountability pressures that might be passed down to teachers. A summary of the main findings of each chapter are presented in *CHAPTER 6*, organised by the research questions that frame the study. Finally, in the concluding chapter, *CHAPTER 7*, key empirical observations of the study are highlighted, limitations and further research ideas are presented and recommendations are presented. The thesis ends with some short final thoughts.

COMPENDIUM OF PUBLICATIONS

- Natalie Browes & Hülya K Altinyelken (2019). The instrumentation of test-based accountability in the autonomous Dutch system. *Journal of Education Policy*, 1-22. https://doi.org/10.1080/02680939.2019.1689577

- Natalie Browes & Antoni Verger (2020). Global governance through peer review: the Dutch experience of OECD reviews of National Policies for Education. *Critical Policy Studies*, 1-20. https://doi.org/10.1080/19460171.2020.1865177

- Natalie Browes & Hülya K Altinyelken (forth). Professionalism in the Era of Accountability: Role Discrepancy & Working Strategies amongst Teachers in the Netherlands. Under review at *British Journal of Sociology.*

- Natalie Browes (2021). Test-based Accountability & Perceived Pressure in an Autonomous Education System: Does School Performance Affect Teacher Experience? *Educational Assessment, Evaluation and Accountability.* https://doi.org/10.1007/s11092-021-09365-9

REFERENCES

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, *18*(2), 215–228.

Ball, S. J. (2016). Neoliberal education? Confronting the slouching beast. *Policy Futures in Education*, *14*(8).

Braun, A., & Maguire, M. (2018). Critical Studies in Education Doing without believing – enacting policy in the English primary school. *Critical Studies in Education*, *00*(00), 1–15.

Breakspear, S. (2012), "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance", *OECD Education Working Papers*, No. 71, OECD Publishing, Paris

Camphuijsen, M. K. (2020). Coping with performance expectations: towards a deeper understanding of variation in school principals' responses to accountability demands. *Educational Assessment, Evaluation and Accountability*, 1-27.

Coleman, J., Scheider, B., Plank, S., Shiller, K. S., Shouse, R., Wong, H., & Lee, S. (Eds.). (1997). Redesigning American education. Boulder, CO: Westview Press.

Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research*, *37*(8), 677–692.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers college record*, *106*(6), 1145-1176.

Deterding, N. M., & Waters, M. C. (2018). Flexible coding of in-depth interviews: A twenty-first-century approach. Sociological Methods & Research.

Evers, J. & Kneyber, R. (2015). *Flip the System.* Published in Association with Education International, Routledge.

Evetts, J. (2009). New Professionalism and New Public Management: Changes, Continuities and Consequences. *Comparative Sociology*, *8*(2), 247–266.

Eurostat. (2019). Hours of Work, Annual Statistics. Last accessed on 19th October 2020. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Hours_of_work_-_annual_statistics

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

Holloway, J. & Brass, J. (2018) Making accountable teachers: the terrors and pleasures of performativity, Journal of Education Policy, 33:3, 361-382

Hanushek, E., Link, S. & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, *104*, pp.212-232.

Hanushek, E., Raymond, M. (2004) The Effect of School Accountability Systems on the Level and Distribution of Student Achievement. *Journal of the European Economic Association, Volume 2, Issue 2-3,* 406–415.

Hargreaves, A. (2000). Teachers and Teaching: Theory and Practice and Professional Learning Four Ages of Professionalism and Professional Learning. *Teachers and Teaching: History and Practice*, *6*(2), 37–41.

Helliwell, J., Layard, R., & Sachs, J. (2019). World Happiness Report 2019, New York: Sustainable Development Solutions Network.

Karsten, S. (1999). Neoliberal Education Reform in the Netherlands. *Comparative Education*, *35*(3), 303–317.

Ko, J., Cheng, Y.C. and Lee, T.T.H. (2016). The development of school autonomy and accountability in Hong Kong: multiple changes in governance, work, curriculum, and learning. *International Journal of Educational Management*, *30*(7).

Le Galès, P. (2010). Policy instruments and governance. *The SAGE Handbook of Governance*, 142159.

Luhmann, N., 1990, "Essays on Self-Reference", Columbia University Press, New York.

Maroy, C., & Van Zanten, A. (2009). Regulation and competition among schools in six European localities. *Sociologie du travail*, *51*, e67-e79.

Mintrop, H. (2003). The limits of sanctions in low-performing schools. *Education policy analysis archives*, *11*, 3.

Mintrop, H. (2007). School Improvement Under Test-Driven Accountability : A Comparison of High- and Low-Performing Middle Schools in California CSE Report 717 Heinrich Mintrop National Center for Research on Evaluation , Standards , and Student Testing (CRESST) / Universit, 1522(310).

Nathaniel, P., Pendergast, L. L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states. *Teaching and Teacher Education*, *59*, 492-502.

Nichols, S. L., Glass, G. V, & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14(1).*

Nusche, D Braun, H., Gábor, H., & Santiago, P. (2014), OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing. http://dx.doi.org/10.1787/9789264211940-en

OECD (2011). School autonomy and accountability: Are they related to student performance? *PISA in Focus*. Available at: http://www.oecd.org/pisa/pisaproducts/pisainfocus/48910490.pdf

OECD (2013). Chapter 4: School Governance, Assessments and Accountability *in What Makes Schools Successful? Resources, Policies and Practices – Volume IV*

OECD. (2016). Reviews of national policies for education: The Netherlands, foundations for the future.

OECD. (2017). *Education at a Glance 2017: OECD Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/eag-2017-en.

Pagès, M. (2021). Enacting performance-based accountability in a Southern European school system: between administrative and market logics. *Educational Assessment, Evaluation and Accountability*, 1-27.

Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods: Integrating Theory and Practice* (4th edition). Thousand Oaks, CA: Sage.

Pawson, R., & Tilley, N. (1997). An introduction to scientific realist evaluation. *Evaluation for the 21st century: A handbook*, *1997*, 405-418.

Penninckx, M., Vanhoof, J., Maeyer, S. De, & Petegem, P. Van. (2016). Research Papers in Education Enquiry into the side effects of school inspection in a 'low-stakes' inspection context, *1522*(March)

Porter, T., & Webb, M. (2008). The role of the OECD in the orchestration of global knowledge networks. *The OECD and transnational governance*, 43-59.

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. sage.

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), The handbook of global education policy (pp. 128–144). West Sussex: Wiley-Blackwell.

Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. London: Sage.

Schuller, T. (2005). Constructing international policy research: The role of CERI/OECD. European Educational Research Journal, 4(3), 170–180.

Steiner-Khamsi, G. (2014). Cross-national policy borrowing: Understanding reception and translation. *Asia Pacific Journal of Education*, *34*(2), 153-167.

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

Takase, M., Maude, P., & Manias, E. (2006). Role discrepancy: Is it a common problem among nurses? *Journal of Advanced Nursing*, *54*(6), 751–759.

Tatto, M. T. (2007). Reforming teaching globally. United Kingdom: Cambridge University Press

Taylor, I. (2007). Discretion and Control in Education: The Teacher as Street-level Bureaucrat. Educational Management Administration &Leadership 35(4): 555-571.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of mixed methods research*, *1*(1), 77-100.

Thiel, C., & 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).

UNESCO & Global Monitoring Report. (2017). Accountability in Education: Meeting our Commitments. *Policy Brief*, 1–505.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, *44*(3), 519-558.

Verger, A. (2014). Why Do Policy-Makers Adopt Global Education Policies? Toward a Research Framework on the Varying Role of Ideas in Education Reform. Current Issues in Comparative Education, 16(2), 14-29.

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

Wilkins, C. (2011). Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system. *Professional Development in Education, 37:3, 389-409*

Woessmann, L., Lüdemann, E., Schütz, G., & West, M. (2007), "School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003", *OECD Education Working Papers*, No. 13, OECD Publishing, Paris

World Bank: Arcia, G., Macdonald, K., Patrinos, H. A., & Porta, E. (2011). School Autonomy and Accountability. System Assessment and Benchmarking for Education Results, (April), 1–29.

World Bank. (2016). Education Global Practice, Smarter Education Systems for Brighter Futures, SABER School Autonomy and Accountability. World Bank Group Files. Available at: http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/brief/SABER_SAA_Brief.pdf

Yin, R. K. (2009). Case study research: Design and methods (4th Ed.). Thousand Oaks, CA: Sage.

CHAPTER 2

A GLOBAL POLICY IN A NATIONAL CONTEXT:

Test-based Accountability in the Autonomous Dutch System

This chapter is a pre-print version of: Natalie Browes & Hülya K Altinyelken (2019). The instrumentation of test-based accountability in the autonomous Dutch system, *Journal of Education Policy*, 1-22. Published online on November 17th 2019, in Journal of Education Policy by Taylor and Francis. DOI: 10.1080/02680939.2019.1689577

INTRODUCTION

In recent years, there has been a considerable and growing interest in the role of accountability in education: the instruments through which it is realised, the actors it involves and the effects that it has. What is particularly noteworthy is the apparent similarities in accountability policies being adopted by quite different education systems across the globe. With standardised testing and data playing a central role, accountability is often synonymous with test-based accountability or 'TBA.' Moreover, often accountability does not come in isolation but alongside or in close pursuit of policies that afford greater autonomy and responsibility at the school level. This is based on the theory that by increasing autonomy (both administrative and educational), schools can adapt teaching to the local context, become more innovative and responsive and enjoy greater community engagement while increasing teachers' motivation and feelings of ownership (OECD 2013). In this context, strengthening accountability thus becomes a way for governments to maintain quality from a distance.

Despite research showing mixed results at best (see for example Hanushek, Link & Woessmann 2013 or Ko, Cheng & Lee 2016), this 'school autonomy and (test-based) accountability' phenomenon, promoted by international organisations such as the Organisation for Economic Cooperation and Development (OECD) and the World Bank (Rizvi & Lingard 2010) has spread worldwide. It has its roots in neoliberal Anglo-Saxon countries, is shaped by the New Public management values of efficiency and effectiveness and is achieved through mechanisms of decentralisation, standardisation

and marketization. Such reforms have had far-reaching implications. A 'datafication' of the profession (see for example, Lewis & Holloway 2018) is changing what education quality means, what is valued, and ultimately, changing teaching practices and professional identities (Day 2002; Evetts 2003; Evetts 2009; Hargreaves 2000). This wider reform package, having now spread well beyond its Anglo-Saxon roots, has been coined the global education reform movement or 'GERM' by influential academics such as Hargreaves and Sahlberg. It has led to the conceptualisation of a global education policy space (Verger, Novelli & Altinyelken 2012) and has resulted in the dramatic expansion of research into areas such as 'global policy transfer' and 'policy borrowing' (Steiner-Khamsi 2014; Verger 2014).

This paper, based on expert interviews and reviews of key policy documents, looks at the influence of the GERM and the drivers behind TBA in one country – the Netherlands. Focusing on the primary school level, this is done by adopting an instrument-centred approach to better understand policy selection and the development of TBA, examining the role of school autonomy within this. The research forms part of a larger, comparative project, known as *ReformEd*, which examines the evolution and enactment of school autonomy and accountability policies in a number of contexts (reformedproject.eu).

Perhaps due to its noteworthy characteristics (a highly decentralised system with a long history of inspection and testing), school autonomy and accountability in the Netherlands has received international attention (OECD 2016b; UNESCO & GMR 2017; World Bank 2011). Yet these discussions remain largely normative, lacking a broader, critical perspective on how and why these policies have developed. While in many ways, the reforms in the Netherlands are typically characteristic of the GERM (increased autonomy followed by greater accountability through the testing of core learning standards), in other ways, the Dutch system is unique. Indeed, while school autonomy increased in the late 20th century it was certainly not born in this period (see Eurydice 2008). Rather, 'Freedom of Education' has been a defining characteristic of the system for over 100 years, and remains crucial in shaping education policy today. Therefore, in the age of school autonomy *with* accountability, this case provides an insight into the tensions that can also exist between these two elements and highlights the absolute importance of national context, institutional legacy and the domestic policy agenda in 'global policy' research.

The paper seeks to answer three main questions: (Q1) How and why has test-based accountability developed in the Dutch education system over recent years? (Q2) What were the drivers behind the selection of TBA and why were seemingly 'interventionist' policies adopted in such an autonomous system? (Q3) What effect has Freedom of Education had on the shape that TBA has taken, and with what impacts?

Following this introduction, the theoretical framework used to analyse the development of the GERM and TBA is presented with a focus on the dynamics of policy borrowing and policy instrumentation. An

explanation of the research methods and the national context are then provided. Results are then presented in three sub-sections, each addressing one of the three research questions outlined above. Finally, in the discussion and conclusion, findings are more broadly situated in theory to show how they can be used to better understand and guide policy reform and instrument choice, and the implications for further research are discussed. The paper argues that the core principles of the GERM are clearly present in the Netherlands, and that (certainly for a period) TBA developed incrementally, advancing towards higher stakes for schools and transforming in unforeseen ways. Evidence shows the importance of the international environment in the Dutch adoption of TBA, particularly the OECD and its Programme for International Student Assessment (PISA) which acted as both an influencer and legitimiser of GERM-based policies. At the same time, findings point to the central role of local institutions and the domestic policy agenda in shaping instrument selection and adaption. The Dutch tradition of 'Freedom of Education' has been found to have played a particularly important role in mitigating TBA.

ADOPTING A SYSTEMIC POLICY INSTRUMENTS APPROACH TO TBA

Sahlberg (2016) highlights three main principles of the GERM, namely *decentralisation* (providing schools with expanded decision-making powers), *standards* (detailing what students are expected to know and encouraging focused and fluid learning), and *accountability*, 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).

The standardised test ties these three principles together (Ball, Junemann & Santori 2017; Verger, Parcerisa, & Fontdevila 2018). These tests tend to focus on what has become accepted internationally (partly as a consequence of the OECD's PISA initiative) as the core learning areas: numeracy and literacy. Student scores – or what are misleadingly referred to as 'outcomes' – in relation to relative or absolute standards, become synonymous with education quality. To encourage transparency and market dynamics, these scores are usually published. In many ways, TBA is an effective policy instrument: the concept of school quality is simplified, standardised and quantified, meaning it can be assessed from a distance and encouraged through school competition. Verger et al (2018) have succinctly summarised these main GERM principles and the role of standardised testing in each case (table 2).

Table 2: GERM Principles & the Role of National Assessments

GERM principle	Definition and main policies	Role of national assessments	
Standards	Prescription of a national curriculum and establishment of quality standards	National assessments used to make sure schools meet and adhere to evaluable learning standards	
Decentralization	Transfer of competences and authority from the central government to lower administrative levels	National assessments used to control state, regional, provincial and local authorities	
	Devolution of managerial and/or pedagogical responsibilities to principals and schools	National assessments used to govern at a distance a range of autonomous providers through the principles of outcomes-based management	
Accountability	Educational actors made responsible for their actions/results through some form of evaluation linked to consequences	*Administrative accountability*	Test results attached to incentives or sanctions for schools, principals and teachers
		Market accountability	Test results used to inform school choice and promote school competition

Verger, Parcerisa & Fontdevila (2019) p. 5

The majority of literature focuses on the application and impact of these three principles in Anglo-Saxon, neoliberal contexts. These contexts have almost become synonymous with the GERM and are considered to have adopted particularly high-stakes forms of TBA in the pursuit of strengthening school choice and market mechanisms (Verger et al 2018). Policy reform in the US, beginning over 15 years ago with ‘No Child Left behind’ and evolving into ‘Every Student Succeeds’ has received particular attention. The reforms intended to increase school quality through market competition and by setting clear student achievement targets, with rewards (such as teacher bonuses) and sanctions (such as school closure) attached to student outcomes. Promises of autonomy have led to the increase of publicly-funded, (semi-)autonomous schools, such as charter schools in the US and academies in England. These schools are particularly subject to narrow, output-based forms of accountability, which brings into question whether this ‘autonomy’ is in fact experienced by educators (Crawford 2001; Kauko & Salokangas 2015). Indeed, with such high-stakes attached to test scores, research has shown time and again that teachers feel under great pressure to perform within increasingly narrow boundaries (see Ball 2003), and constricted and overworked as a result (for example, Berryhill, Linney & Fromewick 2009; Day & Smethem 2009). It has also been repeatedly found that such policies produce undesired practices including; teaching to the test, reshaping the test pool, educational triage and cheating (see de Wolf & Janssens 2007 for an overview). Wider, more secondary impacts of these issues include increasing inequalities between schools and students and a narrowing and instrumentalization of education, as well as teacher deprofessionalisation and job dissatisfaction (Evers & Kneyber 2015; Verger & Parcerisa

2017). Yet despite these negative effects, the implementation of prescribed learning standards, standardised tests and accountability, continue to grow (Verger et al 2018). Intergovernmental organisations – particularly the OECD – have played a key role here, promoting these policies as 'best practice' and a way to improve student learning outcomes, performance in international assessments such as PISA, and ultimately, economic competitiveness (Rizvi and Lingard 2010).

A body of research has investigated the growth and spread of such policies through the lens of policy-borrowing. The work of Steiner-Khamsi (see for example 2013, 2014 & 2016) is particularly significant here, emphasising the need to take a systems theory approach to understand the cross-national policy borrowing phenomenon. This, it is argued, bridges the gap between a macro, neo-institutionalist approach – in which important contextual differences that exist between 'global' policies are essentially glossed over – and those country-case approaches adopted by cultural anthropology in which the homogenising effects of globalisation are downplayed or even 'denounced' (Steiner-Khamsi 2014, p. 161). Indeed, the strong similarities between the instruments used to govern education systems around the world cannot be ignored, yet at the same time we see important differences between the characteristics of these instruments including the techniques and tools that comprise them and their final uses and impacts. It is thus essential to recognise the dynamics behind the *reception* – 'the political, economic, and cultural reasons that account for the attractiveness of a reform from elsewhere' and the *translation* – 'the act of local adaptation, modification, or re-framing' of imported reforms (Steiner-Khamsi 2014, p.153). These processes and the contextually-specific drivers behind them result in a uniquely-evolved set of policy instruments and tools. Indeed, the Netherlands has quite different educational and administrative traditions than those Anglo-Saxon 'GERM pioneers' previously discussed. It straddles the border between such neoliberal models and the more centralised and welfare-based continental and Scandinavian models (Karsten 1999). This context plays a crucial role in (re-)shaping what, at least superficially, appear to be the same policies. Significantly, governments do not import foreign policies blindly, but select those elements that fit their own agenda, and may make strategic use of 'internationally-approved' instruments as a way to justify and legitimise national reform, in a process referred to as *externalisation* (Schriewer 1990; Steiner-Khamsi 2014 & 2016).

This paper couples a policy borrowing and a policy instruments approach to understand the development of TBA in the Netherlands. This approach recognises the national and international drivers behind reform, but zooms in on the policy instruments selected to achieve this reform and the evolution of these instruments. By unpacking the logics of choice (Le Galès 2010), the policy instruments approach seeks to overcome the shortcomings of the functionalist approach to policy adoption, which assumes that instruments are natural or neutral, and freely available for selection by policy-makers (Kassim & le Gales 2010). This functionalist approach has been criticised as a gross oversimplification of the policy process (Kassim & le Gales 2010; Peters 2002). Policy-makers do not freely choose the most

technically-adept instrument in light of policy goals, rather, selection is a far more complex, contingent and bounded process. As well as guided by perceived policy effectiveness or 'instrumentality' (Capano & Lippi 2016), instrument choice is also guided by acceptance. Acceptance is based on perceptions of appropriateness or 'legitimacy' (Capano & Lippi 2016), which in turn, are mediated by a range of factors, such as institutions, interests, ideas, individuals and the international environment (Peters 2002) as well as by decision-makers' preferences (Capano & Lippi 2016).

Further, a policy instruments perspective offers important theories on how adopted instruments and selected techniques and tools might develop. Understanding policy instruments as life forms, or 'institutions' (Lascoumes & Le Gales 2007), independent of the decisions that created them, helps us understand their evolution, diffusion, and impact. Predicting their path and their final outcomes can therefore be difficult (Bezes 2007). Finally, as instruments are not equal but favour some actors over others, it can also be challenging to predict those who will benefit from the instrument and those who will be disadvantaged (Lascoumes & Le Gales 2007; Kassim & Le Gales 2010).

By taking an approach that recognises the drivers behind instrument adoption, and the factors that shape it, importance is given to the complex, multi-level mechanisms at work, and the interaction between different spaces. This complements the notion of a 'global policy space' and theories on cross-national policy borrowing (Verger et al 2012). Indeed, by acknowledging the influencing and legitimising processes at the international level, while simultaneously recognising the importance of national context and domestic agenda, we can better understand the growth of TBA and the spread of the GERM. Given the significance of school autonomy in the Netherlands, much of the discourse around instrument and tool choice is ideological. Therefore, it is not always the most 'straight forward' or the most effective policy that is adopted.

METHODS

As noted, this paper forms part of a wider study that takes the Netherlands as one case through which to explore autonomy and accountability policies. This case-study approach enables at once a recognition of nation-bound policy and the institutional legacy that shapes it, while acknowledging those influencing factors that sit outside of these boundaries (Yin 2009).

Findings are based on two main sources. Firstly, from 25 interviews conducted between October 2017 and February 2018. Key organisations were identified that play an important role in educational policy formation, testing and accountability, and participants were chosen by their position in these organisations. Once initial stakeholders were identified, a snowball sampling method (Teddlie and Yu 2007) was used to find other participants and organisations should any have been overlooked. Participants included policy-makers, policy-consultants, policy-designers, members of advisory

councils, councils of school boards, managers of school boards, representatives of teacher, parent and student organisations, representatives of national testing and curriculum organisations, and academics. A small number of these also had a second (or more accurately, first) role as teachers, and one as a school principal. Their dual roles helped to enrich the data by also providing a more practice-based insight into TBA. Apart from in four cases, where two colleagues were interviewed simultaneously, all interviews were one-to-one. In one instance, due to participant preference, key questions were emailed to two respondents who sent a joint written reply. As such, these respondents are cited together. The total number of participants totalled 31. For anonymity purposes, quoted participants have been described broadly.

As this study forms part of a comparative project, an interview protocol was developed amongst the project team, with interviews taking a semi-structured form to respect contextual differences and participant expertise. The protocol concentrated primarily on the process of policy formation and the drivers behind policy selection. This included the problems that policies aimed to address, how these problems were legitimised and by who, as well as the sources or 'inspiration' of solutions. It also focused on policy consensus, negotiation and adaptation.

Interviews were conducted in English (with Dutch used when necessary), were audio recorded and lasted between 45 and 110 minutes. Interviews were later transcribed, coded and analysed using the programme Atlas.ti. Codes were developed in line with the interview protocol and included for example 'problem content,' 'problem definer' and 'problem consensus.' Codes also carefully reflected the considerations taken into account during the policy formation process, such as 'pol consid_ideological' or 'pol consid_technical' as well as reflecting the changes policies went through, e.g. 'pol_negotiation.' A code book was developed amongst the project team in which codes were defined and clarification provided on when they should and should not be used. For more information on the interview protocol and code book, see Fontdevila (forthcoming). Using these codes more specific, national-level drivers were then identified through the use of memos. These included, for example, 'PISA scores,' 'poor student competencies' and 'Freedom of Education.' In this way, codes adopted a critical focus that uncovered dynamic negotiation processes, the drivers behind instrument and tool selection and policy evolution.

The second source of information was online policy documents. Documents connected to key legislation, as outlined in section five were predominately found on the websites of the Ministry of Education, Science and Culture, and the website of the House of Representatives. These documents consisted of proposals of law, explanatory reports, parliamentary responses, advice from consultative bodies, letters from the Minister of Education to the House of Representatives, and the final publications of law. In keeping with the interview protocol, when reading these documents particular attention was paid to the way in which problems and solutions were framed, and how they reflected the processes of

policy negotiation and evolution. The documents provided important background information which helped to corroborate the information provided in interviews and deepen understanding of TBA policies in general.

NATIONAL CONTEXT

Dutch education is a highly autonomous, highly tracked system (OECD 2016a). Based on performance at primary level, students attend various forms of mainstream secondary education; 'VMBO' (leading to senior vocational training after graduation at 16), 'HAVO' (leading to university of applied science after graduation at 17) and 'VWO' (leading to scientific university after graduation at 18). There are also tracks within these three streams, making the Dutch system the most tracked amongst OECD economies (OECD 2014). The current structure of education and the policy dynamics that shape it must be understood by going back to key events and periods of reform over the last 100 years

Freedom of Education

In the early 1900's, Dutch society was pillarized along religious lines. Decades of struggle, in which Christian groups called for equal government funding for their own schools, culminated in the 'Pacification of 1917.' This saw an amendment to Article 23 of the Constitution, known as 'Freedom of Education.' The Act affords three fundamental freedoms to *all* schools (provided key criteria are met); 'freedom of school establishment', 'freedom of direction' (shaping schools around a particular religious or philosophical belief) and 'freedom of organisation' (including; choice of teaching methods, materials, and personnel) (de Groof & Glenn 2007). The Act also states that the overall quality of education should be a government concern. The Act resulted in the private establishment of many schools across the country. These independent or 'bijzonder' schools quickly came to outnumber public schools. Rather than being managed by local government they were managed by school boards, often composed of parents, and had the right to set their own selection policies and to receive extra funding from private sources.

Decentralisation

Decades later, during a period of decentralisation that began in the 1980s under a Conservative (confessional-neoliberal) government (Sleegers & Wesselingh 1995), differences between public and independent schools were reduced. A crucial step came in 1994 with the signing of the Schevenings Beraad: an agreement between central government, school boards, teacher unions and parent organisations. Municipalities gained core responsibilities from central government (becoming providers of school housing and 'educational-disadvantages'), but at the same time handed over management-powers of public schools to independent bodies. Public schools therefore became (administratively) very similar to their 'independent' counterparts. Importantly, the Schevenings Beraad also laid out a new system of funding. Rather than a claims-based system, school boards would be given a single 'lump-

sum' payment, calculated on a per-capita basis.[4] This resulted in boards and schools having greater decision-making powers over matters such as staff employment and (theoretically) over educational tools and content (Karsten 1999). However, the new system also loaded them with great administrative and financial responsibilities. 'Lump-sum funding' was introduced sector by sector until it finally came to primary-schools in 2006. These momentous reforms imitated those happening elsewhere, led by conservative neoliberal governments in the UK and the US, and driven by a neoliberal push for efficiency (reducing government spending and administrative burdens) rather than by ideological notions of school autonomy.

These developments have led to a system where almost all schools in the Netherlands are publicly-funded yet independently-managed. They have also resulted in a highly diverse system where parents are free to choose between 'general' schools, schools with a particular pedagogical foundation and/or schools that adopt a particular religious or philosophical approach. The Dutch system can therefore be described as de facto 'quasi-market' (see Bartlett 1993 in van Zanten 2009). Yet, while these freedoms have meant diversity and choice they have also led to concerns over segregation and inequality, particularly in large urban centres with high immigrant populations (Altinyelken & Karsten 2015; Vedder 2006).

The Governance of Dutch Schools

It is important to clarify that while school autonomy is a distinguishing feature of the Dutch system, officially, this autonomy lies with the school boards. Although this distinction may have been insignificant in the early-mid 20th century (when boards were generally parent-run and served one school), more recent neoliberal reforms resulted in mergers of boards and their increasing professionalisation. Now, in many cases (particularly in urban areas) one board is responsible for several schools. Recent data reveals that over 90% of decisions, both curricular and administrative, are made at the level of the school board, or – given that boards may devolve some or many of these responsibilities to school management – at a lower level (OECD 2016a). The direct role of the central government in schools is thus officially rather limited. As well as having overall responsibility for the teacher training system, the government's main role is as quality assurer: to ensure that minimum requirements are met in terms of school management, student care and student attainment. Concerning this third element, the government is responsible for the setting of core learning standards and national tests.

Testing has traditionally occurred at two key stages: at the end of primary and the end of secondary education. In their final year of primary (mostly aged 12), the majority of students have customarily sat

[4] Students deemed to have certain social disadvantages or learning difficulties would receive extra funding.

an ‘end-test,’ made by the public-private ‘CITO’ organisation.[5] This test was first used in 1968 as a meritocratic, class-independent way of streaming students into secondary education. Since the 1980’s, test-based ‘student monitoring systems’ (abbreviated to ‘LVS’) have also been used by the majority of primary schools. Also developed by CITO and designed primarily for formative purposes, they are based on biannual tests covering the core learning areas, taken throughout primary school.

With regard to learning standards, given the constitutionally-confined role of the government in educational practices, the implementation of any form of core curricula has been challenging. The result of decades of increasing input regulation from the 1970’s was a messy and overcrowded curriculum (Kuiper & Berkvens 2013) leaving schools confused about what was most worth teaching and concerned over excessive government involvement. To address this, the number of objectives has since been reduced and de-specified, resulting in extremely broad goals that offer little guidance. This is also problematic for schools. It has led to the development of the ‘TULE’: a much more detailed guide as to how students can achieve these broad goals. This in turn has been incorporated into the textbooks of many of the big publishers. Yet, within a high-stakes testing environment (which will be presented next), many schools are keen to follow these textbooks carefully, bringing into question how much curriculum autonomy is actually experienced in schools.

The following findings section presents the formation and evolution of TBA in the Netherlands, examining the multi-level drivers behind policy selection and in particular, the impact of the Dutch legacy of educational autonomy. The section is divided into three, each subsection addressing one of the research questions listed in the introduction of the paper.

FINDINGS: THE DEVELOPMENT & EVOLUTION OF TBA IN THE NETHERLANDS

The Development of TBA

The Netherlands has not been immune to the GERM. Over recent decades the core principles of this global reform have been progressively assimilated into the Dutch education system. With decentralisation policies handing greater decision-making powers to boards and schools, culminating in 2006 with the lump-sum funding of primary schools, the government began taking measures to ensure a firmer grip on education quality. Several techniques and tools were developed to establish, monitor and stimulate student attainment in core areas. While these operate in different ways and through

[5] Since 2004, the functions of the CITO organisation have been divided between its public and private branches. The primary end-test is government-commissioned and its development thus falls under the public branch. The private branch of the organisation develops and sells testing products and services for education and business customers over the world. This includes the LVS tests in the Netherlands.

different actors, they all function through data generated by standardised tests. They are therefore discussed collectively as 'TBA'. The section below reveals the incremental development of TBA at the primary school level, with the primary end-test expanding quite beyond its original streaming function, and advancing towards higher and higher stakes for schools and boards. This will be illustrated through five key moments:

Testing linked to inspection regime: In-keeping with the autonomous system, Dutch school inspection has worked on a 'proportional' basis since 2002 (de Wolf, Verkroost & Franssen 2017). Whereas the intensity of inspection was initially based on school self-evaluation, in 2007 it became risk-based and outcome-centred. At the primary level, this meant that schools performing below average in the end-test for three consecutive years would be considered 'at risk' and likely to receive an on-site inspection. Based on this, schools deemed low-performing would be labelled '(very) weak'. As well as resulting in more intrusive monitoring, these labels could be harmful for attracting new students.

Interviews revealed that risk-based inspection models were heavily influenced by similar systems in the US and aimed to increase efficiency (more targeted inspection) and effectiveness (increasing incentives for school improvement). The model simultaneously fulfilled more ideological criteria, leaving adequately performing schools to their own devices. This method of inspection placed a new significance on testing. It became not only high-stakes for students, but for schools. We can understand it both as driven *by* and a driver *of* a 'datafication' of education:

> In this period, a lot of data became available for schools in the Netherlands so this enabled the Inspectorate to build risk models and to have a smarter way of inspection; (…) to focus on the underperforming schools and leave alone, or trust, the better performing schools. Because, especially the better performing schools, said 'the inspector is of no use to us anymore, we copied their way of looking at quality, we can benchmark from the internet now as all the information from the Inspectorate is made publicly available…
>
> (Member of the Inspectorate)

More recently, inspection has shifted to the level of the school board, yet results-based risk-monitoring at the school level remains an important part of the framework. Further, as part of a government push for 'excellence' in education, the Inspectorate can now also award schools with labels of 'good' and 'excellent.'

Sanctions attached to test results: In the late 2000's there were important discussions concerning *who* was ultimately responsible for school quality and *how* this quality could be ensured. In 2009, these discussions culminated with the passing of the 'Good Education, Good Governance Act'. The Act reminded school boards of their responsibility for school quality and required them to comply with a good governance code. Significantly, the Act also introduced sanctions for underperforming schools. Schools that continuously performed below average in the end-test could effectively be closed by the Minister of Education. In such a traditionally autonomous system, the Act can be considered somewhat

momentous, and has been described as ‘unprecedented’ (Waslander 2010) in terms of the level of government involvement it enabled:

> This means that national government *can* interfere in schools related to education quality, and *that* is unprecedented. Before, because we have this tradition of autonomy and especially freedom of education, we wouldn’t tolerate – ‘we’ - the school boards, the association of school boards – wouldn’t tolerate national government interfering so strongly, so that’s why it’s key.
>
> (Academic 1)

Detailed core standards: In 2010, (when sentiment was for *reduced* input regulation), the government introduced somewhat uncharacteristically-detailed learning standards in the core areas of language and numeracy. In an attempt to raise basic quality and encourage continuous learning paths, the ‘Reference Levels in Dutch Language and Numeracy Act’ specifies the minimum competencies that students should have at key points in their school careers.

> Well, it was the first time that the government was really setting standards. For the levels in numeracy and literacy to be reached by schools…by law…that was quite new.
>
> (Policy-maker 1)

To ensure that schools were meeting these standards at the primary level, they would be formally evaluated through the long-standing primary end-test.

The publication of test results: Over the last decade, the Inspectorate’s ‘quality card’ system[6] has become increasingly replaced by the online platform ‘Windows for Accountability,’ or ‘Vensters’. This platform, introduced nationally to secondary-schools in 2010 and to primary-schools a few years later, combines centralised quantitative data (such as student numbers, average end-test scores, and student placement in secondary education), with qualitative school-held data (such as the school’s mission, plan, and annual report). This model of accountability lends itself particularly well to the Netherlands, reinforcing its quasi-market character. Although school quality indicators were originally introduced as a government *reaction* rather than a GERM-inspired policy (following the publication of school rankings in a national newspaper), ‘Vensters’ is now being promoted as a valuable benchmarking and information tool. Visitors to the site are, for example, encouraged to ‘compare schools.’

Testing made compulsory: Standardised testing became a priority of the government in the late 2000´s, first under a centrist (neoliberal-labour-confessional) coalition, and later taken up with vigour by a Conservative (neoliberal-confessional) coalition. Policy documents from this period emphasise the importance of testing in strengthening core skills, encouraging excellence, and realising continuous learning paths. At the primary level, the introduction of these measures came through the 2014 Act ‘Central End-test and Student and Education Monitoring System’. The language and numeracy reference

[6] Cards outlining the inspector’s main findings.

levels were incorporated into the primary end-test, and for the first time in its history, the test was made compulsory.[7] The Act also made the use of test-based LVS systems compulsory throughout primary education.

While policy documents emphasise the importance of testing for boards, schools, and students, the growing centrality of test-data for purposes of administrative accountability (linked to inspection, sanctions and evaluation of the reference levels), means that it had also become of great importance to the government. LVS test data has also been assumed into the government's push for 'excellence' through achievement-oriented work (opbrengstgericht werken), and by providing national benchmark data, it has become increasingly linked to school accountability (Visscher & Ehren, 2011). Whilst primarily used by school boards, LVS test data (from groups 3, 4 and 6) may also be requested by the Inspectorate should a school be judged 'at risk'.

Conditions for Reform & the Reception of Global Ideas

The Dutch adoption of decentralised decision-making with strengthened test-based administrative and market accountability mechanisms, reflects dominant 'global ideas' concerning effective ways to govern education systems. Based on the accounts of several policy experts, the notion that accountability is the inevitable and necessary counterbalance to autonomy, appears to be somewhat naturalised:

> … because of having more autonomy, you also need more accountability. And if you have accountability mechanisms for the Inspectorate but also the horizontal accountability, like benchmarks and Vensters voor Verantwoording etcetera, you're always going to work with outlines or indicators or standards.
>
> (Academic 1)

Almost all interview respondents noted the importance of accountability within Dutch education. In the majority of cases, this was in reference to *school board* accountability, with the belief that boards should not only be (vertically) accountable to the government and the Inspectorate, but accountable to parents, schools, and society in general.

Yet some of the policies that were adopted to achieve this accountability regime may be considered surprising – seemingly incompatible with the fiercely autonomous context. This includes the development of detailed core standards, the government's ability to close schools, and taking away schools' choice over the testing of their students. To understand this, we must first better understand the domestic policy agenda at the time (Steiner-Khamsi 2014). Crucially, data reveals a widespread concern

[7] This excluded migrant students in the country less than four years, and students expected to go on to (advanced) special education (Ehren & Swanborn 2012).

over the quality of Dutch education in the years prior to the introduction of TBA policies. This concern grew throughout the late 1990's and early 2000's with a sense that Dutch students were no longer mastering the basics of learning. This was largely attributed to a lack of school and board accountability and the government's over-involvement in the organisation of education.[8] Concern over the state of education was soon also taken up by the media and the general public (see Waslander 2010). The government, it was claimed, was failing students. In 2007, under mounting pressure, the government established a parliamentary committee to investigate the state of Dutch education.

In early 2008, the committee (known as 'Dijsselbloem' after its chairman), concluded that there was a general downward trend in the quality of Dutch education. An important justification for this conclusion was the decline and/or stagnation of students' scores in international comparative assessments. PISA scores were a particular focus, with the report emphasising a decline in results (both in absolute and relative terms) and a lack of top performers. The report claimed that the government had intervened too much in the organisation of schools while ignoring its core role. It recommended that the government formulate clearer learning standards and strengthen testing (Dijsselbloem 2008).

Many researchers involved in the enquiry did not stand by its conclusions (Waslander 2010). Overall, stakeholders had given positive accounts of the education system, national periodic tests showed no deterioration, and performance in international tests still remained strong (see Scheerens 2016). Yet the report was a powerful tool: it tapped into public interest and used seemingly 'unbiased' international data to substantiate quality concerns. The result was a perceived need for immediate reform:

> For the Netherlands it was very special that there was a general feeling of urgency. This feeling was in society, education and politics. All [stakeholders] agreed about the problem and the wish to fix it quick.
>
> (Curriculum policy-designers 1&2)

This paved the way for the quick introduction of TBA tools that, at any other time, may have been viewed as overly interventionist. This is evident in one policy-maker's account of the passing of the Good Education, Good Governance Act:

> …In the end, the discussion was 'we cannot allow, as a society, that we have schools where young people are educated for years and this education is not [good] quality. And also, you [the government] have the responsibility for this [quality]. So in the end this law had a broad majority; it was supported broadly in Parliament.
>
> (Policy-maker 2)

[8] In particular, this alludes to a government reform known as the 'Studiehuis' which restructured student learning at the upper secondary level.

In the weeks and months that followed the publication of the Dijsselbloem report, core education legislation was sent to the House of Representatives: TBA had been found to be the solution to the problems facing Dutch education. The role of international discourse in shaping and legitimising national policy solutions, appears to have been significant. The introduction of clearer standards and compulsory testing, promoted by influential organisations such as the OECD and through expertly-branded tools such as PISA, was (is) seen as an effective and efficient way to improve system quality. The influence of these normative views in the framing of policy solutions are reflected in Dijsselbloem's test-based recommendations and throughout government discourse at the time. The following citation; taken from a letter sent by the (then) Minister and Secretaries of Education to the House of Representatives just weeks after Dijsselbloem's recommendations, is nicely illustrative of this and demonstrative of the process of 'externalisation.'

> International comparative research shows that central examinations contribute to educational quality. Education systems characterized by central examinations and autonomy for schools score better on the internationally comparable PISA math test (see for example, Woessman 2005). Research also shows that transparency of learning achievements is key to specific educational policy, both at a school and at a national level. Working with reference levels encourages the results-orientation of schools and leads to an improvement in the performance of pupils.
>
> (De Minister & Secretarissen van Onderwijs, Cultuur en Wetenschap, 2008)

Importantly, these internationally-promoted, test-based instruments were, by and large, also nationally appropriate. Given that tests and exams fall under the responsibility of the government, such reform was ideologically acceptable. Moreover, at the primary level the introduction of these measures was a relatively straight-forward process, due largely to the fact that testing structures were already in place (according to policy documents, around 85% of primary schools were already implementing the CITO end-test). This well-established test therefore acted as a convenient vehicle on which to attach new policy tools. The ideological appeal of a more outcomes-based system was acknowledged in several expert interviews, illustrated here by one academic:

> All the input had changed into an idea of measuring the output. And that's an economic way of thinking, but [also] we don't have to change the whole idea of our Freedom of Education... 'it doesn't matter how you teach it if the results are ok'.
>
> (Academic 2)

Seen as both effective and appropriate, from 2010-2014, testing came to dominate Dutch education policy:

> The idea was, and this is one of the things that you should know, is back then in our Ministry of Education, most people said 'if you *really* want to change something in education, you have to introduce a test'.

(Policy-maker 3)

The Translation of TBA in an Autonomous System

The previous section has largely examined what has been referred to as the moment of 'reception': the meeting of global policy ideas with the local context. Yet, in a strongly-autonomous system, where Freedom of Education is closely guarded by both those that make policy as well as those on the receiving end of it, the translation of these ideas can be a difficult negotiation process. The way that policy instruments have developed and evolved in the Netherlands, as in other systems, is unique: adapting to fit national ideologies of legitimacy and advancing and transforming through this into an unexpected form. This final section of findings explores in particular, how the Dutch tradition of Freedom of Education has shaped TBA and with what impacts.

Tensions between central policy and school autonomy are not only apparent throughout the Dutch system but are integral to it. They are seen in the disputed role of the school Inspectorate, resulting in a complex quality-labelling system,[9] as well as in the long (and continuing) balancing-act over curriculum regulation, resulting in vague learning goals. They are also key to understanding the increasing number of mediatory organisations operating between the government and school(board)s, providing an indirect entry point for more process-oriented policies (see Waslander, Hooge & Drewes 2016). Finally, as will now be analysed, these tensions have played a central role in the re-contextualisation of the compulsory standardised testing Act.

As outlined, a decade ago a focus on core competencies, continuous learning paths, and 'excellence' became central to the government's quest for better education. The (then) CITO end-test, and LVS-tests (with the purpose of providing baseline and interim data) were chosen as the vehicles through which to achieve this goal at the primary level, and in January 2012 a proposal was sent to the House of Representatives to make these tests compulsory. The proposal was met with much debate. Concerns were predominately institutionally-based and ideologically-driven. The prescription of one particular end-test for all schools compromised the notion of freedom, and the proposition that this would automatically be *CITO's* test, compromised the notions of fairness. Policy stakeholders feared that once again, the government was interfering in a domain not their own and threatening schools' freedom of organisation. This concern was widespread, held by various parties along the political spectrum and by independent advisory bodies. One teacher union even threatened a boycott of the CITO end-test.

True to the Dutch system, the result was a political compromise, and significant amendments were made to the Act. To reduce infringement on freedom of organisation, schools would be able to *choose* their

[9] Recent parliamentary debates have resulted in changes to the Inspectorate's role in issuing quality labels. Schools can *request* a 'good' or 'excellent' label, but cannot be *automatically awarded* one, as this controversially places the Inspectorate as the adjudicator of quality education.

testing and monitoring instruments. This led to a liberalisation of the primary end-test. The government would now be responsible for the old CITO end-test (now known as the 'central end-test'), and private companies could bid to become test providers, so long as key quality and content criteria were met. Schools would have free choice regarding which test to use, with the costs publicly subsidised.

Interestingly, even prior to these negotiations, the 'compulsory testing Act' proposal was itself a product of compromise, having already incorporated significant modifications to the nature of the end-test. A concern in parliament and in education more broadly over the rise of a 'testing-culture,' coupled with fears of a 'checkout culture'[10] resulted in a proposal to push back the date of the test from February to April. This meant it would be taken once students had already applied to secondary education. This had the effect of making the teacher's advice more important for student streaming, and the end-test less high-stakes for students.[11] It has also further removed these tests from their original function.

These compromises have led to a number of unintended effects. Since the passing of the compulsory testing Act in 2014, data has revealed rising student inequalities as a result of teacher bias (Inspectorate of Education 2018 p.22). Socially-disadvantaged students it seems, are less likely to be given 'the benefit of the doubt' when receiving advice for their secondary school placement. Further, a recent survey of over 2000 primary school teachers revealed the significant parental pressures they face to give favourable advice (CNV Onderwijs & EenVandaag 2018), including numerous reports of threatening behaviour and even law suits.

To list another 'side-effect', the number of end-test providers has been growing year on year since the test's market liberalisation in 2014. In 2018, these providers numbered six.[12] This has undermined the test's comparability function. For secondary schools admitting new students, it has led to the question: can students who take different end-tests be directly compared? For the government and the Inspectorate, it has led to the equally difficult question: can schools that use different end-tests be directly compared? These questions are currently occupying policy. In December 2016, the report 'comparability of end-tests' was published, which compared secondary-school advice given by the various tests (only three at the time). Findings showed there were differences between test advice that could not be explained by region or student background (Emons Glas & Berding-Oldersma 2016).

The tensions and struggles that exist between school autonomy and accountability in the Dutch system can therefore result in oddly-shaped and less technically-effective policies. The implications of this were a particular concern for one policy-maker:

[10] Students and teachers no longer applying themselves for the remaining school year following the test.

[11] Given that this advice is often based on LVS-test results from the previous years, the reduced importance of one test has led to the increased importance of others.

[12] This dropped to five this year due to insufficient order numbers for one provider.

> At this moment we are nibbling at the safeguards we have to make this [a] balanced system. My personal opinion is that we are going the wrong way with this, so we should not throw these end-tests in primary education on the market. That's my strong opinion.
>
> (Policy-maker 3)

Yet, whilst at least at the policy-level, Freedom of Education might be sometimes seen as an obstruction, it is viewed by others as having a protective role, offering some degree of defence against the GERM. This view was most notably expressed by those policy experts who also held teaching positions:

> I think a lot of the accountability-based measures in Holland were actually quite late compared to other countries. So if you look at the UK or the US, they started doing accountability-based output models far earlier and we're quite late actually. And it reared its head [here] but it didn't take as much hold on our education system as it did in other countries. And that's mainly due to the Freedom of Education Act.
>
> (Policy-advisor/teacher 1)

Indeed, the role that school autonomy has played in mitigating the GERM is significant. While the GERM undoubtedly arrived in the Netherlands and, certainly for a period, saw TBA advance rapidly, it has developed in a less extreme way than in those 'pioneer' systems previously described. Performance-based pay and value-added modelling initiatives were dropped at their pilot stages, the Inspectorate's quality indicators have recently become broader again, and, of course, along with compulsory testing at primary level came a choice of tests and lowered (end-test) stakes for students. Further, while standards in language and numeracy are detailed in a relative sense, it is certainly not accurate to speak of a prescribed national curriculum. TBA techniques have also been enforced with less vigour than elsewhere: despite the 'Good Education, Good Governance' Act attaching results to funding, policy experts claimed that only in one case has a school been closed as a result of poor performance. Institutional autonomy, it appears, has resulted in a softer GERM.

DISCUSSION

By tracing the journey of TBA in one country and examining the drivers behind policy selection, this paper reflects on the complex, contextual and contingent nature of the national adoption of a 'global' policy. In many ways, the GERM has spread into the Dutch education system as it has into many countries around the world, evidenced through the presence of its core principles: decentralisation, standards, and accountability. As part of this wider reform, TBA evolved incrementally with its expansion proceeding in a somewhat ad hoc manner and the stakes attached to standardised tests increasing for boards and for schools.

Yet, understanding TBA in the Netherlands simply as an imported policy does not reflect the dynamic, multilevel processes at play, nor the important contextual drivers. As policy-borrowing theorists suggest, governments do not import policies unless they suit their domestic agenda, and do not import entire packages but select those elements that suit this agenda. In the early 21st century, a widespread belief that the Dutch education system was deteriorating and concern over what was perceived to be unchecked school(board) autonomy provided the widespread receptiveness and necessary push for increased accountability. These public concerns had been shaped and legitimised through the government's careful manipulation of national and international data. The sense of reform urgency was strong enough to enable the selection of policies that otherwise may be considered overly intrusive for the Dutch context. Unlike in the Anglo-Saxon cases therefore, in the Netherlands TBA was not driven by market mechanisms and school choice (a tradition that long preceded the reform), but rather by this perceived need to introduce a counter-balance to school and board autonomy that had increased a decade earlier (see also Verger et al 2018).

It is important to situate policy instrument choice in the Netherlands within the wider arena and to acknowledge the significance of internationally-prominent ideas and promoted practices. Given the repeated referral to OECD ideas and data throughout interviewee accounts and national policy documents, it may be reasonable to conclude that the organisation played an important influencing role in national policy selection. Moreover, through the process of externalisation (Schriewer 1990; Steiner-Khamsi 2014), the status of these instruments and tools as technically-effective and internationally approved, was used by the Dutch government to legitimise TBA reform. Vitally, OECD data and discourse were therefore used to frame both *policy problems* (declining PISA scores) as well as *policy solutions* (standards and testing).

Findings have clearly indicated however that perceived effectiveness is not the sole driver of instrument choice. Indeed, the extent to which instruments are considered contextually appropriate is highly significant (Capano & Lippi 2017). Perceptions of appropriateness are in turn, heavily shaped by institutional history (Peters 2002). In the Netherlands, where policy is heavily influenced by constitutional school freedoms, legitimacy seeking is an essential yet often challenging task.

In several ways, TBA is harmonious with the autonomous Dutch system. It was ideologically appropriate, considering standards and testing are a constitutional responsibility of the government and an output focus (theoretically) respects schools' rights to organise how learning takes place. It was also practically appropriate, considering the longstanding presence of the primary end-test in the majority of schools. However, whilst pre-established tests were a convenient vehicle for new accountability tools, it is important to highlight the significant functional change of these tests as a result. The primary end-test (introduced as an equality measure in the late 1960's) and the LVS-tests (introduced as a student-learning aide in the 1980's) have now assumed a core function of school accountability.

Despite the ways in which TBA aligned with the Dutch system, the instrument and its component parts also underwent important changes during the process of translation. For the most part, the local adaption of TBA was the result of a careful and necessary negotiation through the complexities of school autonomy. It is on this point that we must understand the relationship between autonomy and accountability in a second way: not as complementary but as conflicting. At times, these tensions resulted in a loss of instrumentality in favour of legitimacy (Capano & Lippi 2016). The principle on which TBA is hinged – the standardisation of testing, enabling the direct comparison of schools – was compromised. Allowances made for freedom of organisation resulted in a liberalisation of the test and a growing number of test providers. This illustrates the significant and often mitigating effect that a traditionally highly autonomous system has had on a global education policy. This is essential for understanding the differences that exist between TBA in the Netherlands and elsewhere. Certainly, while the Netherlands has adopted the main GERM principles, in reality the tools and consequences attached to these principles are softer than in many other systems. It also advocates the importance of adopting a 'socio-historic' approach to understanding policy instruments, such as that proposed by Verger, Fontdevila & Parcerisa (2019).

The adaptation of the end-test in this way also lends credence to the notion of policy instruments as life forms: growing and morphing independently of the decisions that created them. In the Netherlands, this has helped fuel a budding testing industry, both in terms of the companies developing compulsory tests and the accoutrements associated with such tests (re-focused textbooks, test preparation tools, online platforms that organise and analyse test data…). Further, many of the impacts of TBA have been unintended and perhaps to a large part, unforeseen. Whilst the enactment of TBA is outside the scope of this paper, complementary research has indicated that rather worryingly, certain students have been disserved by the compulsory testing act. In part, this is a result of the unequal advice generated across tests (Emons Glas & Berding-Oldersma 2016), with interview accounts (supported by online data) revealing that many schools have switched providers under the belief that the CITO-designed end-test is linguistically complex, and less child-friendly than others. Perhaps more worryingly, these impacts also include a rise in student inequalities, following the primacy given to teacher advice in the streaming of students.

CONCLUSIONS

This paper has presented a critical view of policy selection, formulation and evolution, and has illustrated the importance of adopting a case-centred but not case-contained understanding of 'global policy'. The in-depth focus has revealed the complexities and challenges that exist in a system that is simultaneously influenced by external ideas and moulded by its own strong institutional traditions. These challenges can be categorised on two levels: challenges for policy-makers and challenges for

policy-enactors. With regard to the former, experts are well aware of the difficulties surrounding the adoption of accountability policy in a highly autonomous system. Constitutional freedoms, while offering some degree of protection to schools, teachers and boards against government intervention, also make it more challenging to adopt and implement effective policies. In turn, this can also make it difficult for the government to ensure a firm grasp on issues of quality and equity.

In terms of the challenges facing policy-enactors, this paper has suggested that, in education systems with high levels of school autonomy, high-stakes accountability risks jeopardising professional freedoms. That is, in an environment of standardisation, where there exists a narrow, output-based view of school quality, teachers may in reality feel constricted by accountability – weary of exercising their curricular and pedagogical autonomy and focused instead on those tested competencies. This may be the direct effect of the bureaucratic accountability that has been a focus of this paper, but equally may be the (more indirect) result of school-board and market pressures. Given the important, dialectical relationship between education reform and teacher professionalism (see for example Helgøy & Homme 2007), further research into the enactment of TBA and the effects it has on (theoretically autonomous) teaching professionals is warranted here.

Not only in the Netherlands but internationally, the drivers behind policy are complex, dynamic, and often conflicting. Governments face the difficult task of achieving a balanced system: one where education quality can be (centrally) assured while school and educator autonomy is truly respected. By exerting pressure on schools to conform to limited interpretations of quality, policy-makers are jeopardising this balance.

REFERENCES

Altinyelken, H. K., & Karsten, S. (2015). The Netherlands: structure, policies, controversies. In T. Corner (Ed.), Education in the European Union: pre-2003 Member States (pp. 305-324). (Education around the world). London: Bloomsbury Academic.

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.

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.

Bezes, P (2007). ‘The Hidden Politics of Administrative Reform’, *Governance* 20(1): 23–56.

Bovens, M. (2007). Analyzing and assessing accountability: A conceptual framework. European Law Journal, 13(4), 447–468.

CNV Onderwijs & EenVandaag (2018). *Onderzoek: 'Leraren onder druk om schooladvies.'* Available at:https://eenvandaag.avrotros.nl/fileadmin/user_upload/PDF/Rapportage_Peiling_Leraren_Schooladvies.pdf

Crawford, J. R. (2001). Teacher Autonomy and Accountability in Charter Schools. Education and Urban Society, 33(2), 186–200.

Day, C. (2002). "School reform and transitions in teacher professionalism and identity". *International journal of educational research*, *37*(8), 677-692.

Day, C. & Smethem, L. (2009). The effects of reform: Have teachers really lost their sense of professionalism? *Journal of Educational Change*, *10*(2–3), 141–157.

Dijsselbloem, J. (2008). Tijd voor onderwijs [Time for education]. Eindrapport parlementair onderzoek onderwijsvernieuwingen [Final report of the parliamentary research commission on educational reform]. The Hague: House of Representatives of the Netherlands

Ehren, M. & Swanborn, M. (2012). Policy and Practice Strategic data use of schools in accountability systems in *School Effectiveness and School Improvement : An International Journal of Research*, (April 2013), 37–41.

Emons, W., Glas, C., Berding-Oldersma, P. (2016). *Rapportage Vergelijkbaarheid Eindtoetsen.* Available at:https://www.eerstekamer.nl/overig/20170126/rapportage_vergelijkbaarheid/f=y.pdf

Eurydice, (2008). *School Autonomy in Europe, Policies and Measures.* Available at: https://publications.europa.eu/en/publication-detail/-/publication/102bb131-8105-4599-9367-377946471af3/language-en

Evers, J. & Kneyber, R. (2015). *Flip the System.* Published in Association with Education International, Routledge.

Evetts, J. (2003). "The sociological analysis of professionalism. Occupational change in the modern world". *International Sociology*, 18(2): 395-415.

Evetts, J. (2009). "New Professionalism and New Public Management: Changes, Continuities and Consequences". *Comparative Sociology*, 8(2): 247-266.

Fontdevila, C. (forthcoming). *Researching the adoption of School Autonomy with Accountability reforms. A methodological note on country case studies.* Reformed Methodological Papers. Number 1. Available at: http://reformedproject.eu/tools/

de Groof, J. & Glenn, L. (2007). *Balancing freedom, autonomy, and accountability in education.* WolfPublishers

Hanushek, E.A., Link, S. and Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, *104*, pp.212-232.

Hargreaves, A. (2000). "Teachers and Teaching: Theory and Practice and Professional Learning Four Ages of Professionalism and Professional Learning". *Teachers and Teaching: History and Practice*, 6(2): 37-41.

Helgøy, I., & Homme, A. (2007). Towards a New Professionalism in School? A Comparative Study of Teacher Autonomy in Norway and Sweden. European Educational Research Journal, 6(3), 232–249.

Inspectorate of Education, Ministry of Education, Culture & Science (2018). Summary report: *The state of education 2016/17* (Eng.)

Karsten, S., (1999). Neoliberal Education Reform in the Netherlands. *Comparative Education, 35* (3) 303-317

Kassim, H., & Le Galès, P. (2010). Exploring governance in a multi-level polity: A policy instruments approach. *West European Politics*, *33*(1), 1–2

Kauko, J., & Salokangas, M. (2015). The evaluation and steering of English academy schools through inspection and examinations: national visions and local practices. British Educational Research Journal, 41(6), 1108–1124.

Ko, J., Cheng, Y.C. and Lee, T.T.H. (2016). The development of school autonomy and accountability in Hong Kong: multiple changes in governance, work, curriculum, and learning. *International Journal of Educational Management*, *30*(7).

Kuiper, W., & Berkvens, J. (2013). *Balancing curriculum regulation and freedom across Europe*

Lascoumes, P., & Le Gales, P. (2007). Understanding Public Policy Through its Instruments: From the Nature of Instruments to the Sociology of Public Policy. *Governance*, *20*(1), 1–21

Le Galès, P. (2010). Policy instruments and governance. *The SAGE Handbook of Governance*, 142–159.

Lewis, S., & Holloway, J. (2018). Datafying the teaching "profession": remaking the professional teacher in the image of data. *Cambridge Journal of Education*, (March), 1–17.

de Minister en Secretarissen van Onderwijs, Cultuur en Wetenschap, (2008). (31 332) *Doorlopende Leerlijnen Taal en Rekenen: Brief van de minister en staatssecretarissen van Onderwijs, Cultuur en Wetenschap* (Nr.3)

OECD (2013). Chapter 4: School Governance, Assessments and Accountability *in What Makes Schools Successful? Resources, Policies and Practices – Volume IV*

OECD (2014). Education Policy Outlook, Netherlands http://www.oecd.org/education/EDUCATION%20POLICY%20OUTLOOK_NETHERLANDS_EN%20.pdf

OECD (2016a). Index of school autonomy, school characteristics and science performance: Results based on school principals' reports in *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*, PISA, OECD Publishing, Paris.

OECD (2016b). *Reviews of national policies for education: The Netherlands, foundations for the future*.

Peters, B. (2002). The politics of tool choice. In L. M. Salomon (Ed.), *The tools of government. A guide to the new governance,* pp. 364–402. New York: Oxford University Press.

Rizvi, F. and Lingard, B. (2010). Globalising education policy. New York: Routledge.

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The handbook of global education policy* (pp.128–144). West Sussex: Wiley-Blackwell.

Scheerens, J. (2016). The Netherlands: The intersection of international achievement testing and educational policy development. In L. Volante (Ed.), *The intersection of international achievement testing and educational policy: global perspectives on large-scale reform* (pp. 108-121). New York, NY: Routledge.

Schriewer, J. (1990). The method of comparison and the need for externalization: Methodological criteria and sociological concepts. In J. Schriewer & B. Holmes (Eds.), *Theories and methods in comparative education* (pp. 3–52). Bern: Lang.

Seegers, P. & Wesselingh, A. (1995). Decentralisation, professionalism and autonomy NL sleegers1995, *47*(2), 199–207.

Steiner-Khamsi, G. (2013). What is wrong with the "what-went-right" approach in educational policy? *European Educational Research Journal*.

Steiner-Khamsi, G. (2014). Cross-national policy borrowing: understanding reception and translation, Asia Pacific Journal of Education, 34:2, 153-167

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

Teddlie, C., & F. Yu. (2007). "Mixed Methods Sampling: A Typology With Examples." *Journal of Mixed Methods Research* 1 (1): 77-100.

UNESCO, & Global Monitoring Report. (2017). Accountability in Education: Meeting our Commitments. *Policy Brief*, 1–505.

Van Zanten*, A. (2009). Competitive arenas and schools' logics of action: a European comparison. *Compare: A Journal of Comparative and International Education*, *39*(1), 85–98.

Vedder, P. (2006). Black and White Schools in the Netherlands, *European Education*, 38:2, 36-49

Verger, A. (2014). Why Do Policy-Makers Adopt Global Education Policies? Toward a Research Framework on the Varying Role of Ideas in Education Reform. *Current Issues in Comparative Education*, *16*(2), 14-29.

Verger, A., Novelli, M., & Altinyelken, H. K. (2012). Global education policy and international development: An introductory framework. *Global education policy and international development: New agendas, issues and policies*, 3-32.

Verger, A., & Parcerisa, L. (2017). A Difficult Relationship: Accountability Policies and Teachers. International Evidence and Key Premises For Future Research. In M. Akiba & G. LeTendre (eds.). International Handbook of Teacher Quality and Policy (pp.241-254). New York: Routledge.

Verger, A., Parcerisa, L., & Fontdevila, C. (2018). The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. Educational Review, 1–26.

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

Visscher, A., & Ehren, M. (2011). De eenvoud en complexiteit van Opbrengstgericht Werken, 1–39.

Waslander, S. (2010). Government, School Autonomy, and Legitimacy: Why the Dutch Government is Adopting an Unprecedented Level of Interference with Independent Schools. *Journal of School Choice*, *4*(4), 398–417.

Waslander, S., Hooge, E., & Drewes, T. (2016). Steering Dynamics in the Dutch Education System. *European Journal of Education*.

Woessman, L. (2005) The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-REPEAT and PISA, *in Education Economics*, vol.13(2), p.143-169.

de Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education*, *33*(3), 379–396.

de Wolf, I., Verkroost, J., Franssen, H. (2017). Supervision and Accountability in the Dutch System, In *The Dutch Way in Education pp. 111-121.* Onderwijs Maak je Samen & De Brink Foundation

Yin, R. (2009) Case study research, design and method. 4. London: Sage Publications

CHAPTER 3

SAWA & GLOBAL GOVERNANCE:
The Dutch Experience of OECD Reviews of National Policies for Education

This chapter is a pre-print version of: Natalie Browes & Antoni Verger (2020). Global Governance through Peer Review: The Dutch Experience of OECD Reviews of National Policies for Education, *Critical Policy Studies*. Published online on December 23rd 2020, in Critical Policy Studies by Taylor and Francis. DOI: 10.1080/19460171.2020.1865177

INTRODUCTION

A growing body of research is taking an interest in the influence that intergovernmental organisations, such as the Organisation for Economic Co-operation and Development (OECD), exert in the national policy domain. Whilst the OECD has no formal policy role in its member countries, its presence and sway within the national policy arena is significant. So far, the majority of research has turned to the OECD's gathering and use of data to exercise this influence (Grek 2014). In the case of education, the OECD's 'Programme for International Student Assessment' (PISA) has been massively successful in this regard, with league tables creating competition between nations and providing clear education goals to strive towards (Sellar & Lingard 2013; Niemann, Martens & Teltemann 2017). This has resulted in the intensification of policy learning and transfer – often between systems which, administratively speaking, are quite different – and has led to the significant restructuring of many education systems, including those previously regarded as well performing (Fischman, Topper, Silova, Goebel & Holloway 2019; Waldow & Steiner-Khamsi 2019). This has been attributed to the fact that more and more governments and key stakeholders conceive student achievement in international learning assessments as a benchmark of educational quality and a driver of economic success and attractiveness (Komatsu & Rappleye 2019).

However, PISA is not the only means the OECD has for promoting policy ideas and change. This paper focuses on an under-researched governance mechanism at the OECD's command, that of the policy

review: a peer review performed upon request from national (or regional) governments, that provides an overview of the country's main sectoral policies and challenges, with recommendations and guidance on how to improve the functioning and the performance of a system. In the education sector, the OECD policy review instrument most widely used is the 'Reviews of National Policies for Education' (RNPE), which focuses on multiple dimensions of education systems.[13] Since 1998, an average of 2.4 RNPE have been conducted per year.[14] Initially, these were conducted in OECD member countries, but are now increasingly conducted in non-member countries also. RNPE can be characterised by their interactive nature, their reliance on field visits, and the length of the policy cycle they involve - lasting up to 12 months or more (Grek, 2017). According to the OECD:

> Education Policy Reviews are based on an in-depth analysis of a country's strengths and weaknesses, using various sources of available data such as OECD Programme for International Student Assessment, Teaching and Learning International Survey and The Survey of Adult Skills, national statistics, research on relevant policies and intensive discussions with stakeholders. They draw on policy experience from benchmarking countries and include expert analysis of the key aspects of education policy and practice being investigated.[15]

This paper examines the RNPE process and its enactment and implications in one OECD member state: the Netherlands. Whilst the Dutch are somewhat prolific participants in OECD initiatives and reviews, this research focuses on the most recent RNPE, published in 2016 (see OECD 2016). In particular, it examines how issues related to school governance, which are at the core of the OECD's education policy agenda (school autonomy with test-based accountability), are framed and advanced through the review, in a country where these governance structures are already in place.

By doing so, it contributes to our understanding of peer reviews as a global governance mechanism in key ways. Firstly, it adds to much needed empirical data in this domain. Given that RNPE have only been analysed in a few national cases (ref), we still lack an understanding of the review process and beyond: how policy influence is manifested throughout, and what national conditions might result in particular (policy) outcomes. Further, the majority of research has thus far shone a critical light on the OECD's use of benchmarking data (PISA in particular) as a mode of transnational governance (Broome & Quirk 2015; Grek 2009; Niemann et al 2017). This study broadens our understanding of the governance tools at the IO's disposal, and shows that in many ways, policy recommendations – simplified and transformed into 'neutral,' 'technical' packages – should equally be conceived and problematised as a form of institutional power. Finally, the research seeks to overcome the blind spots

---

[13] Since 2005, thematic reviews focusing on specific topics (teachers, early childhood education, immigrant education, evaluation policies…) are also part of the OECD review portfolio (Istance 2011)

[14] Source: authors with data from the OECD iLibrary.

[15] See: http://www.oecd.org/education/Flyer%20for%20Education%20Policy%20Advice%20and%20Implementation%202014%20(2).pdf

arising from, on the one hand, globalist perspectives of policy transfer that see IOs as influential actors successfully disseminating their agendas within national policy spaces, and on the other, sceptical views that see national actors instrumentalizing IOs to gain autonomy at the national level and help advance their policy preferences (Martell 2007; Mundy, Green, Lingard & Verger 2016). It asks; why member states might participate in reviews, who is involved in the different stages of the review, how ideas are shared, negotiated and promoted throughout the process, and what are the policy outcomes and impacts of the review.

THE IDEAS GAME: THE ROLE OF POLICY REVIEWS IN THE TRANSNATIONAL GOVERNANCE OF EDUCATION

Policy Reviews as a Form and Practice of Institutional Power

In the absence of legal and/or financial regulatory instruments, 'the OECD is bound to play the so-called ideas game through which it collects, manipulates and diffuses data, knowledge, visions and ideas to its member countries and to a still larger extent, to a series of non-member countries' (Marcussen 2004, 15). In order to play this game efficiently, the OECD operates as an ideational arbitrator and as a policy entrepreneur with several governance instruments at its disposal, such as benchmarking, data-gathering, idea generation, and peer reviewing. Peer reviewing has been characterised as a practice revolving around multilateral surveillance and indirect forms of coercion (Porter & Webb 2008; Schuller 2005). While such reviews are popularly conducted by international organisations (IOs), the OECD has been particularly active in using them (Pagani 2002).

Country reviews are led by a team of reviewers that combines both OECD staff and international independent experts. In the educational sector, the review process works as follows: as a first step, the Ministry of Education (MoE) of the reviewed country and the review team define the review's scope and thematic focus. Afterwards, the MoE tends to commission a background paper that will serve as the basis of the review. After some desk research, the review team visits the country for a period of about 15 days. In this visit, the team meets key education stakeholders and visits a sample of schools and other educational institutions. With the primary and secondary data retrieved, they produce a report draft, whose content is discussed with the MoE before publication. Once the report is ready, it is disseminated through different channels and, ideally, there will be a follow-up stage in which the team gives advice to the MoE regarding the implementation of recommendations.[16] Whilst 'peer review requires standards

[16] See: http://www.oecd.org/education/Flyer%20for%20Education%20Policy%20Advice%20and%20Implementation%202014%20(2).pdf

and criteria against which a member state's policies can be reviewed' (Porter & Webb 2008, 6) these criteria are not always explicit and accordingly, there is some level of discretion.

Through country reviews, national representatives, external experts and the OECD agree on formulating a set of best practices and norms for appropriate behaviour and reflect on what could be the most effective policy instruments to ensure the internationalisation of these practices and norms in the national policy domain (Drezner 2000; Marcussen 2004). Mahon and McBride (2009) consider that the OECD's policy work can be characterised as both an inquisitive and meditative mode of regulation. Through practices such as benchmarking and peer reviews 'member states are not obligated to follow-up specific policies, but they are required to 'open up' to others to examine and critically judge what they are doing' (Mahon & McBride 2009, 6). These practices are apparently more meditative than inquisitive in the sense that they 'are mainly framed as discussion among experts about what is the best way or ways of doing something' (Jacobsson 2006, cited in Mahon & McBride 2009, 6). Nonetheless, despite often being portrayed as a policy learning tool, country reviews also involve indirect forms of coercion and moral pressure. As Marcussen (2004, 18) observes, this is the 'only means that the OECD has at its disposal in its attempt to execute multilateral surveillance.'

All these accounts highlight the normative dimension of IOs' policy work, and the key role of socialisation therein. Indeed, to a great extent, policy reviews are conceived to socialize national representatives with transnational norms, through a process that can be seen as 'inducting actors into the norms and rules of a given community' leading to 'sustained compliance based on the internalisation of these new norms' (Checkel, 2005, 804, in Grek 2017, 2): they are simply accepted as 'the right thing to do' (Meyer, 2005, in Grek 2017, 3).

Through country reviews and other knowledge exchange initiatives, national representatives and other stakeholders do not only learn about international policy trends, but also 'develop their personal and technical skills, and even sometimes develop their personalities and feelings of belonging' (Marcussen 2005, 16).

The Politics of OECD Policy Reviews

In education, OECD policy reviews have received limited scholarly attention, particularly compared to the vast attention that PISA has received. This is unfortunate if we take into account that RNPEs can contribute to diffusion and internalisation of the OECD's educational agenda among national policy-makers, and to consolidate the centrality of the OECD as a policy expert in education (Grek 2017). A recent literature review by Verger, Fontdevila and Parcerisa (2019) on the role of the OECD in disseminating accountability policies and norms in education notes that, in some cases, when picked up by the media, country reviews have the potential to trigger domestic dynamics similar to those sparked by PISA. In countries as different as Chile (Parcerisa & Falabella 2017) and Norway (Møller & Skedsmo

2013; Steiner-Khamsi, Karseth & Baek 2020), RNPE recommendations on accountability have conditioned concrete policy changes (see Verger et al 2019).

However, it is not the norm for OECD reviews to receive significant attention within the national education policy space, and the commissioning of a review is not necessarily the prelude to policy change (Verger et al 2019). The usefulness of reviews 'depends in part on the willingness of the country to confront issues and to be candid in the information it supplies’ (Schuller 2005, 177). Such readiness cannot be taken for granted. The ultimate impact of OECD reviews depends on their reception amongst senior policy-makers, but also amongst the domestic public (Porter & Webb 2008). The OECD is explicit that the effectiveness of peer review relies on peer pressure resulting from informal dialogues, public scrutiny and comparisons, and the impact of public opinion.[17]

An indicator of the influence of OECD knowledge products, such as review reports, is the explicit reference of these products in national political and policy processes. Of course, this does not necessarily mean that anything has been learnt from them or that policy is being changed, but it does at least illustrate a belief that an organisation such as the OECD has a fundamental raison d’être and enjoys legitimacy in a particular policy field (Marcussen 2004; Steiner-Khamsi et al 2020). What is more difficult to capture empirically nonetheless is the normative effect of policy reviews, given the fact that this type of effect is rather indirect and that, overall, normative changes are gradual and difficult to observe.

In this paper, we argue that understanding a country’s motives to commission a review is necessary to subsequently understand the impact and the (potential) influence of the review. As described, literature on global policy studies presents a dichotomy: between globalist accounts that may overinflate the influence of IOs whilst overlooking the agendas of national governments, and sceptical views that do not recognize the influential regulative processes, driven by IOs, that shape these agendas (Martell 2007; Mundy et al 2016). The fact that several studies on OECD reviews in education show that, the recommendations of the review lose centrality once the commissioning government has been replaced (see Halton 2003; Grek 2016), supports, to some extent, the notion of national governments being relatively autonomous from the IO, and commissioning reviews with a political purpose. If not interested in the review’s findings, the government might act as a gatekeeper – limiting its influence by restricting media attention (see Verger et al 2019).

Overall, beyond trying to find out whose interests policy reviews serve the most, it is more appropriate to understand reviews as a process in which both the OECD and the commissioning country might have mutual returns. For the OECD, reviews are not only a way of expanding ideational authority and influence, but also a source of income. Education policy reviews cost between 150,000 to 250,000 Euros,

[17] Cf. https://www.oecd.org/site/peerreview/peerpressurearelatedconcept.htm. See also Pagani (2002).

and they are a significant source of income in a context in which OECD staff, especially in expanding divisions such as the Education and Skills division, are expected to sell more and more services to both member and non-member countries (Lewis, Sellar & Lingard 2016). Furthermore, the fact that national governments may attempt to politically instrumentalize the review does not necessarily place the IO in a situation of subordination. As some authors have observed, the OECD is not only aware of being 'used' by its member countries, but is frequently 'eager to be instrumentalized' (Centeno 2017, 100), as its agenda is only likely to succeed, acquire legitimacy, and enjoy wider circulation if its proposals resonate with national interests (Toledo Silva, 2020)

Nonetheless, the power and ideas game described in this section is contingent to the position that countries occupy in the implicit hierarchy that exists within any IO (Cox & Jacobson, 1973). This hierarchy is defined by countries' economic and political powers, but it also could depend on matters of policy involvement and compliance; i.e. the degree to which countries can be considered 'good students' of the IO. Thus, both the form of power that prevails between the OECD and its member countries and the autonomy of countries in relation to the OECD depend on the position, status and legitimacy that countries have in the stratified system of states that configure the IO.

POLICY CONTEXT: THE OECD POLICY AGENDA & DUTCH EDUCATION

The OECD Agenda for Education

In public sector reform, the OECD is well known as one of the theorizers and international disseminators of New Public Management (NPM) (Pal 2012). Since the 1980s, the OECD, and in particular its Directorate for Public Governance is at the centre of a global policy network that advocates public sector reform amongst member and non-member countries as a way to promote transparency, policy effectiveness and economic efficiency in public services. To advance this agenda, among other measures, the OECD recommends that countries; fragment public services into smaller managerial units, adopt outcomes-based management approaches, and develop professional leadership and external accountability systems.

These general recommendations have penetrated different policy sectors. Nonetheless, the OECD agenda in education is broad and sophisticated, combining a mix of social equity measures and governance reforms *a la* NPM. If we look at the policy recommendations more frequently found in the most relevant knowledge products of the OECD education division, social equity and comprehensive education (avoiding the separation of students into different tracks at early ages) stand out (Bieber & Martens, 2011). Since the release of the first PISA reports, the OECD also theorizes about and promotes

the adoption of an NPM approach to education governance consisting of the development of school autonomy, standardised testing and outcomes-based accountability (see Wöbmann, Lüdemann, Schütz & West 2007; OECD 2011). PISA has been strategic in the dissemination of this policy model, which can be captured under the acronym 'SAWA' (school autonomy with accountability). According to a study that surveyed the 37 OECD country representatives, 29 countries admitted that PISA/OECD recommendations on assessment and accountability have influenced policy reforms at the national level, whereas 11 countries stated the same in relation to school autonomy and choice policy (Breakspear 2012). The under-studied role of policy reviews is even more intriguing in country contexts, particularly in those such as the Netherlands, which are seemingly well-aligned with the NPM and, specifically, the SAWA education agenda.

Education Policy in the Netherlands: The 'Good Student'

A core and long-standing characteristic of the Dutch education system is its early and diverse student streaming system: separating students into various types of secondary education (from more vocationally to more theoretically-focused) based on primary school performance. Another core characteristic is its high levels of school autonomy (Nusche, Braun, Gábor & Santiago 2014). School boards are responsible for school quality while the government ensures system quality, largely through a well-established school inspectorate (see Glenn & De Groof, 2005). Schools have constitutional freedoms, dating back over a century, including the freedoms of establishment, direction, and organisation. This had led to complex governance networks through which the government 'steers' education (Waslander, Hooge & Drewes 2016). Devolution has expanded over recent decades: NPM-driven decentralisation policies saw boards and schools become responsible for managing their own budgets (known as 'lumpsum' funding), and private foundations take over the management of government-established schools. From the late 2000's, with a widespread concern over slipping quality, this was countered with increased school accountability, including the introduction of core learning standards measured by (soon to become) compulsory, stakes-based testing (Browes & Altinyelken 2019). Dutch school accountability measures are both market- and administrative-based. Particularly at the primary level, parents may refer to published, easily-accessible school test results to guide their choice. Schools performing below average in national tests for three consecutive years will be considered 'very weak' and receive (extra) visits from the Inspectorate. Although by law, schools can be closed for repeated poor performance, in reality this is extremely rare.

Debates surface and subside concerning these fundamental aspects of the system. The autonomy-accountability balance can perhaps best be understood as in a constant political flux: pulled between respecting schools' constitutional freedoms and ensuring sufficient quality. One clear example is the role of the 'end-test,' taken in the final year of primary school. Originally introduced as an optional and objective student streaming tool in the 1960's, in the 2000's the test increasingly became used for school

accountability purposes, and in 2014 was made compulsory. Faced with a backlash over intrusion into schools' freedom to organize (one teacher union threatened to boycott the test), the test was also marketized, allowing schools a choice of provider. This legislation, which included compulsory testing in earlier grades, also resulted in widespread concerns of a growing 'testing culture', seen as damaging to students. To help address these concerns, the end-test's role in student tracking was diminished[18] making teachers' advice the primary determinant of student placement in secondary school. These changes have concerned policy-makers (see Browes & Altinyelken 2019), undermining the comparability of testing (Emons, Glas & Berding-Oldersma 2016) and resulting in growing student inequalities due to teachers' bias (Inspectorate of Education 2017). The system of early student tracking has equally seen considerable debate over the decades. While the existence of 'bridge classes' (combined track classes) for the lower secondary level seemed like a practical way to delay the separation of students, these have been decreasingly common in recent years, particularly in light of market pressures (separating levels is seen as more attractive to parents of higher-performing students). A concern has grown amongst various stakeholders that movement between tracks is increasingly difficult (see for example Onderwijsraad 2010).

OECD Policy Reviews in the Netherlands

The Dutch have been somewhat prolific in their participation in policy reviews. Previous to the 2016 RNPE, OECD reviews were published in 2014 (one focusing on evaluation and assessment, another focusing on vocational education and training), 2010 (a review of migrant education), and 2008 (a review of the tertiary education system). The broadest and most comprehensive of these – the 2016 review in question – looks at the Dutch system from pre-school to upper-secondary level. The review was formally commissioned by the (then) Minister of Education (Jet Bussemaker) and State Secretary of Education (Sander Dekker), and focuses on how the Dutch system can be brought from 'good' to 'great' (see OECD 2016). With this in mind, while praising the governance structures that are rooted into the system (decentralised decision making, standardised testing and a strong inspectorate), and its comparatively solid student outcomes, the review makes key recommendations for improvement. Each chapter of the document is based on one of these governance recommendations. Amongst other things, as developed in the findings section, they advocate greater standardisation at the input level, a more pivotal role of national testing, enhanced appraisal and evaluation systems, increased accountability, and further decentralisation of decision making.

[18] This was done by moving the date of the end-test back, until after students had been selected into secondary education.

METHODS

Data for the study was collected from three main sources. First, interviews with key stakeholders from the MoE and OECD review teams; second, content analysis of the 2016 RNPE; and third, analysis of other key documents (including the Minister's invitation letter to the OECD, terms of reference of the review, and the Minister & Secretary's response letter to parliament). To help understand the *response to* and *impact of* the review, online searches of major Dutch media outlets, the MoE's website, and websites of the government's main advisory body and sector organisations were also conducted. This was not done systematically, but to provide useful background information.

Actors who played a significant role in the review process were invited for an interview. Given that this study delves into the details of the RNPE, it was essential that respondents had played a key role, rather than, for example, holding a more distanced, consultative position. This resulted in a small number of selective, highly-rich interviews. In total, five expert interviews were conducted for the study; three respondents from the OECD review team, and two from the Dutch MoE. A third member of the ministry declined to participate.

Interviews were conducted between the end of May 2019 and July 2019. The majority were face-to-face, yet one was conducted through an online platform. Interviews were semi-structured: based on a protocol, but flexible enough to be adapted to participant experience and expertise. The OECD and MoE protocols overlapped considerably so that data could be triangulated, yet differed on a few key questions. Interview themes included: the participant's personal role in the review, the background of the review (and reasons for its commissioning), the review team, the review process, findings of the review, and the reception and impact of the review. Interviews lasted at least 40 minutes, and many over one hour. Interviews were recorded, transcribed, and coded using a coding scheme that primarily emerged from readings and re-readings of the interviews. The RNPE and key documents were coded using the same scheme.

FINDINGS

Findings are organised into four sections by the questions that framed the research. Firstly, (1) the reasons for the MoE's commissioning of the review are addressed, along with how OECD reviewers saw their role and the role of the review. This is followed by (2) an examination of the agenda setting and review process, before (3) the contents of the review are analysed, with a particular focus on SAWA content, highlighted problems, and policy recommendations. Finally, (4) the launch, reception, use, and impact of the review are discussed.

Reasons for Commissioning the Review

The Netherlands' involvement in the 2016 RNPE was largely taken-for-granted. Policy reviews are perceived by MoE respondents as a necessary way to keep an eye on the functioning of the system, combined with a sense that a national review was overdue. The OECD was a desirable partner with which to collaborate: on the one hand it was seen as an expert, able to offer unique insights, a systems overview, and advice based on multi-level, multi-national data, on the other hand its attractiveness lay in its externality and perceived neutrality. Unlike national organisations, ministry officials did not consider the OECD to have invested interests, a political agenda, or pre-formed notions about the Dutch system. As such, the OECD was seen as the most suitable colluder to help the MoE realize its overarching: the pursuit of excellence (see Annex A in OECD 2016).

> We wanted to do something with alignment, so if we all strive for excellence and if you get all the parties aligned and you can have an ambition and a similar frame through which we look at the system and how it works, then we - without trying to overhaul the whole system which takes a lot of energy - we can start by having a more coherent way of trying to improve and go from good to excellent. So who's going to help us by creating that frame? And then we thought - the only party that we could think of that will be enough outside of our system to get to create some distance as well as enough insight into an education system in a developed country... And that's the OECD (Respondent 3, MoE).

Positioning the OECD as an external, independent expert provides considerable advantages: the MoE can use policy reviews to facilitate government agendas (agendas that might themselves be shaped by OECD discourse), by depoliticising certain debates and providing space for a synoptic vision from which the government can begin reforming or restructuring the system. Here, interviews revealed a certain 'selective utilisation' of the OECD. Its insights were required to help fuel the MoE's arsenal, and ultimately help steer policy in certain directions and away from others. Indeed, certain aspects of the Dutch system appear somewhat protected, and not open to debate. In particular, this included the cornerstones of the system – early student tracking and freedom of education – domains seen as historically-complex and difficult to be captured by an external actor.

> So we didn't want a very big story about early tracking in the Netherlands… we thought it wouldn't be very productive because this discussion has been there for a long time already and no political party is committed to postpone tracking, or at least not the majority (Respondent 4, MoE).

OECD actors, on their part, were quite aware of the political motivations behind the commissioning of the RNPE.

> I think the Netherlands clearly has respect for the OECD report, but it's also very good at commissioning them at a time when they want and they think it could really contribute to the policy dialogue (Respondent 2, OECD).

They too, positioned themselves as the 'independent expert' and saw their role as an enabling one: helping to provide evidence-based clarity on issues that had become highly politicised and sensitive at the national level, by drawing on best practice and reframing them into technical, evidence-based channels.

> And we are, of course, a completely independent body, so I think sometimes this is to raise the stakes a little bit in terms of political dialogue, so this is why we were brought in and I think it helped spark some of the policy dialogue on some issues which was needed (Respondent 2, OECD).

Establishing and Negotiating the Agenda and Review Process

Accounts on both sides describe the review process overall as smooth, ideas as a dialogue and decision-making as collaborative. This included decisions on who was involved in the review team.

> Well the OECD has a list of potential candidates, top people that *they* know and they think are capable of performing a review because they have a certain level of understanding – most often in academia or policy. And they asked us, 'Who do you think would be appropriate?' And I said 'I have no person in particular.' I had a few people, Dutch people, who I knew were on the list and who I knew to be very strong… but also again hobbyhorses, so I said 'I don't want them included'… (Respondent 3, MoE).

OECD review team members interviewed had considerable knowledge of the Dutch system, having been either brought up in it themselves, or having worked within it. They were also Dutch speakers. This is an important characteristic of the Dutch review, it automatically inducted the OECD team into the Dutch system and blurred the lines somewhat between the two parties. Further, contrary to protocol,[19] there was reportedly no background country report produced by the MoE prior to the review. Rather, key pieces of relevant national research were collected by the MoE and made available to the OECD.

> … [the MoE was] just feeding it to us, and I've been trying to dig for further research as well, together with my other colleague, but almost everything that we were trying to find was already being handed to us and that was such a big help (Respondent 2, OECD)

A number of important points can be drawn from this. Firstly, it reflects the strong Dutch tradition of system research and reflection. Given that this was deemed sufficient for background information, it also points to the OECD's respect for this nationally-produced research. Finally, the 'insider' element of the OECD review team might be a factor here, contributing to a perception that the Netherlands was a context already somewhat known.

---

[19] See the OECD's process here: http://www.oecd.org/education/Flyer%20for%20Education%20Policy%20Advice%20and%20Implementation%202014%20(2).pdf

Stakeholder involvement during the process was described across the board as comprehensive, even 'rigorous'. While the OECD had a template for desired respondents (actors, consulted in all review contexts), the MoE appears to have been instrumental in contextualising this template. Consulted stakeholders ranged from teachers and teaching assistants to policy-makers and academics, and also included many intermediary bodies (sector organisations) that are so relevant to the Dutch system.

Yet accounts also reveal a degree of friction: the result of some misaligned expectations during the agenda-setting process and of occasionally contrasting perspectives of the Dutch system. Regarding the former, agenda-setting was recounted as a product of negotiation.

> …it's always a question of how much attention do you give to certain topics. Some doubt for example about 'should early childhood education be part of it?' and some people from the ministry said 'yes' and then we thought, well for the Netherlands this is actually very relevant, so 'okay, then it's part of it' (Respondent 2, OECD).

On the one hand, the OECD positioned itself as flexible and there to be instrumentalized. OECD team accounts describe the review as an opportunity for the Dutch ministry to ask questions about its own system, understanding RNPE as tools that governments can shape to their own needs. This led to the frustration of one member of the ministry in particular, who believed that a commissioned expert should assume more of a driving role in the review process.

> I said 'I want to know what *you* think is weird, what strikes *you* in the Dutch system… what do *you* think is not aligned?' and they said… 'well, what do *you* think this is *your* opportunity' and I was like 'I'm not… I've not hired the OECD to feed you with my suppositions or theses or whatever… yet you give it back to me'. So that was… awkward (Respondent 3, MoE).

On the other hand, the OECD's flexibility had its limits. RNPE have pre-established parameters and the OECD was not afraid to reject ideas that ultimately did not fit this model. This 'power of veto' is significant, enabling the organisation to remain in control of the review process.

> …there was a big debate about putting higher education as well in this study, being very honest that would be too much broadening the scope, so we basically declined that part of the offer 'sorry, this is going to be too much' (Respondent 2, OECD).

During the draft-review process, contrasting perspectives of the Dutch system were revealed particularly with respect to early childhood education. Recommendations included; improving quality, increasing care hours, and addressing 'fragmented' provision and access. Female workers, it was claimed, were particularly disadvantaged by the system. Some MoE actors reportedly found this chapter unclear and overly critical. OECD reviewers claimed to have clarified recommendations, yet maintained their position. Both parties noted they had been satisfied with the resolution of these differences.

> They were not too happy about it, because well... it's like a political discussion in the Netherlands… should women go to work or not, should we have childcare, affordable childcare or preschool and stuff like that. So that was a bit of a political thing that they did not really like. But... But we pushed through, we did not back off, so that was good I think. And this was one of the things that I think had most impact on the Minister at the time, because he really cared about this topic (Respondent 1, OECD).

'Problems' and 'Solutions': A focus on SAWA Policies

The Dutch system has a history of school autonomy, centralised testing and school choice, coupled with a strong Inspectorate of Education, reportedly used as a 'reference' inspectorate (Respondent 5, OECD). In short, the Netherlands is seen as a 'good SAWA student.' Significantly therefore, the OECD's agenda is not to introduce *new* SAWA policy instruments, but to (re)calibrate and expand existing ones.

All seven of the report's key recommendations, or parts of them, can be understood as essential elements of the SAWA policy model. This includes an expansion of decentralised decision making – shifting power from school boards to school leaders, as well as an expansion and hardening of accountability tools, through a greater (singular) role of testing in student tracking and school accountability, enhanced school board accountability, improved school self-evaluation and more selective intake and rigorous appraisal systems of teachers and school leaders. Recommendations also encourage clearer standards; including the development of a national curriculum for early childhood education and teacher competency standards. Finally, there is an emphasis on encouraging top student performers – comparatively speaking, there is almost always room for improvement: 'The Netherlands has more 15-year-old top-performers in basic skills than most of Europe, but is still behind some Asian countries...' (OECD 2016, paragraph heading, 81).

According to OECD reviewers, one of the main issues identified in the Dutch system was a 'SAWA imbalance,' with Dutch accountability tools in particular need of recalibration.

> I mean, New Zealand and the Netherlands probably have the schools with the greatest levels of autonomy… so of course there were issues around, in particular the capacity of the school boards to exercise their autonomy. […] And also, whether that autonomy is well-balanced enough, with enough accountability (Respondent 5, OECD).

OECD respondents expressed a concern that 'freedom of education' was restricting the government's ability to ensure quality and equity, and that accountability mechanisms had become too 'soft' and inefficient. This point was often raised in discussion of the primary end-test, whose liberalisation in 2014 to respect schools' freedom of organisation, dramatically undermined its comparability function.

> ... All the indicators of accountability are becoming more loose over time. I mean, we used to have, at least *the idea* is to have freedom, balanced by accountabilities, so it's checks and balances. But we are

> losing the balance here, because the freedom is getting bigger and bigger, and the checks and balances are getting weaker and weaker (Respondent 1, OECD).

A distinct advantage of autonomy, accountability and evaluation recommendations appears to be their versatility. Delayed student tracking for example, is framed as an equity issue, with the review highlighting 'large performance differences' between tracks as a particular cause for concern. The solution offered is a stronger role of standardised testing:

> An objective track decision requires a single national end of primary test, which could be extended to examine a broader range of competences than at present. Nationally set objective standards on the required scores for each track level should be established and should determine entry to different tracks. Local discretion by primary teachers and the receiving secondary schools create both inconsistency and bias and should be removed from the decision. The transparency of such a system would be fair to all students (OECD 2016, 73)

Not only is standardised testing strategically placed at the centre of the response to (perceived) key weaknesses, but, once this has been done, the scope and function of such tools can then be expanded.

SAWA is not only promoted through the review's core recommendations and the problematisation of particular national policies, but also through the explicit use of policy learning. Benchmarking is a key tool, used by the OECD to put pressure on governments through comparison. This method is employed throughout the review: in its provision of examples of 'good practice' from across the globe, in its use of graphs that rank the Netherlands against other OECD countries in various aspects of education, and – as seen – by drawing on PISA performance to show where gains can be made.

The concept of reference societies – considered to be those systems from which the Netherlands can learn most – is another key aspect of the benchmarking mechanism. In OECD interviews, reference societies provided were often countries that had made significant changes to their systems following a recent review. As well as 'good examples' the notion of *counter reference societies* also surfaced: countries which could seemingly provide no useful policy lessons to the good Dutch student. Reportedly, this was due to different socio-geographical structures and systems of education, but equally, due to perceived poor performance.

> …Well let's be honest, the Netherlands does quite well, so you immediately start looking at some of the countries that are doing really well. So you don't look to countries such as Mexico, which arguably aren't doing so well in terms of education, but [also] in terms of geographical spread and scope [are] so different. So you immediately think of countries such as Scotland, and at the time even Wales started to come up as well as an interesting country. Estonia, Finland, Austria, Denmark, countries that the Netherlands *could* learn from and draw from in terms of their own reflections (Respondent 2, OECD).

Use and Impact of the Review

MoE respondents reported the efforts made by the OECD to maximize the review's visibility and impact upon launch. This was interpreted by one ministry official (Respondent 3) as a 'willingness to help governments improve their education systems.' Yet, such publications are undoubtedly also crucial to the OECD, providing purpose and raising its profile and influence: 'everything goes public, it's the first and main rule' (Respondent 5, OECD). An online search reveals that the published review was picked up by many of the mainstream Dutch newspapers as well as the sector organisations. Takeaway messages vary considerably, with newspaper headlines as diverse as:

> *Little Wrong with the Dutch Education System* (De Telegraaf, 25th May 2016)

to

> *OECD: Dutch Students Unmotivated, Little order in Class, a lot of Talent Unused* (de Volkskrant, 26th May, 2016)

This underscores the (political) adaptability and instrumentality of RNPE. Yet, such headlines were fairly short lived: there was not an extensive, high-profile fall-out from the review, rather, despite the apparent efforts of the OECD, its (direct) impacts appear quite limited. One exception is in the area of early childhood education.

> …so this government has invested maybe 70 million or something in early childhood education and childcare, to have some more intensive… to raise the hours from 10 to 16 hours and also to invest more in quality. But I think there, there was also some political support and also the advice from the OECD so it fits well together (Respondent 4, MoE).

Limited direct impact of the review may largely be due to the nature of its findings, not delivering the insights the MoE desired.

> They [the OECD] did put things on the agenda which I'm sure will pop up or have popped up again here and there. But this was not the great game changer that I hoped it would be (Respondent 3, MoE).

Recommendations that *were* made were not well-aligned with the political environment – requiring a restructuring of the system's governance and intervention into those long-standing, politically-charged domains described earlier in this paper. Instead, in a formal latter responding to the review's findings, the Minister and Education Secretary defended the aspects of the system highlighted as in need of improvement, such as early student tracking 'we indicate that it is undesirable that pupils are not given the opportunity to follow the education that best suits their talents' (MoE, 2016). One OECD team member, aware of the more critical elements of the review, wondered if its launch had been downplayed.

> I think on that day, two other reports came out from the ministry and the message from, I think Sander Dekker, that more money was going to secondary teachers. So basically that buffered the media attention a little bit for the report, and I think we lost media attention because of that. So that's a little bit unfortunate, or potentially it might have been deliberate from the ministry, you never know, as there was some critical parts of the report… (Respondent 2, OECD).

Given the significant changes made to the accountability system in the period preceding the review, there was a sense that pushing government for further changes at the time would not have only garnered frustrations. Rather, different issues were occupying policy debates, in particular, teacher action, demanding improved salaries and reduced work pressures.

Insights into the Dutch institutional context also offer important insights into the review's limited impact. In contrast to what Grek (2017) observed in Sweden, in the Netherlands, the OECD is just one of many respected voices: considerable amounts of research is conducted nationally, and the government prides itself on its consensus-building approach to decision making. RNPE and other such reviews are resources that may help *reframe* but do not *lead* the education debate. Reportedly, changes made to early childhood education were only implemented once further research had first been conducted at the national level:

> …in some EDPC[20] countries, the OECD is really seen as an objective referee or as the grand authority, if they say 'you need to step up' they say, 'oh of course that's what we'll do.' This is not the role or the status of the OECD in the Netherlands. (Respondent 3, MoE).

This is not to say that such reviews lack political use or influence. RNPE can be used to help stifle undesirable social or political sentiments. In the Dutch case, it was felt that the report helped counter growing 'anti-testing' feelings in the country at the time, a sentiment several MoE actors worried would further marginalize the role of the standardised test: 'I think the OECD proved a valuable counterweight to that' (Respondent 3, MoE).

Findings can also be used for future political ammo, perhaps during periods of greater political openness.

> I think it was quite well received, it touched upon certain points for improvement and it helps us in our work. If we have strategic projects then it helps to be able to refer to this report… (Respondent 4, MoE).

Browsing the ministry's website, the report was also found to have been used for retrospective purposes – to justify earlier reforms. Explanations of the school 'lumpsum funding' system for example, are justified using the OECD SAWA model of governance - that increased autonomy leads to improved student performance – and referenced with a link to information on the RNPE.

---

[20] the OECD's 'education policy committee'

Four years have now passed since the 2016 review's publication, offering some time to better observe policy developments and uses of the report. In this time, the Netherlands has ushered in a new government and a new Minister and Secretary of Education. Debates that at the time had subsided, have now reawakened, with movements seen in the direction of OECD recommendations, including proposed changes to the end-test[21] and a collective of influential stakeholders pushing for delayed student tracking.[22] So far, use of the RNPE in these initiatives appears limited, but importantly, the review captured and to some extent contributed to reinforce a climate of opinion that problematised, for example, the Dutch legacy with early tracking.

DISCUSSION & CONCLUSIONS

This study has analysed policy reviews as a mechanism of global governance, and has aimed to better understand the influence of the OECD in the national education sphere through the use of these products. To do so, the paper draws on the Netherlands as a case study where a Review of National Policies for Education, carried out by the OECD, was published in 2016. The research has examined the policy review process in detail through its different stages: from its inception, to the data gathering process, the report's elaboration, its release, and impact.

Findings indicate a departure from previous perspectives. They do not align with globalist, nor with sceptical views of policy transfer, rather, indicate a certain symbiosis. Certainly in the case of the Netherlands, the RNPE should be understood as the product of two parties – the OECD and the national government – pursuing particular self-interests. On the one hand, the MoE was inducted into normative notions of 'best practice' through subtle mechanisms such as socialisation, networking and continuous negotiation. In the review itself, these notions were dressed as scientific, apolitical recommendations, much in the same way as benchmarking data has been used (Broome & Quirk 2015). Yet, the Dutch government commissioned the RNPE for specific reasons, and used the review strategically and selectively. By approbating the OECD's reputation as an independent education expert, issues that had become too sensitive, too contentious or too political at the domestic level, could be redirected into more 'technical' and 'impartial' channels. Rather than intentionally used to legitimate specific reforms or policies, peer reviews are desirable to help reframe the national education debate. This subtle steering mechanism may be particularly useful in decentralised contexts such as the Netherlands, where the government has restricted power over schools and boards. As such, in the absence of coercive instruments at its disposal, the institutional power of the OECD in the global governance scenario relies,

[21] https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/wetsvoorstel-doorstroomtoetsen-po last accessed on 12 June 2020
[22] https://toekomstvanonsonderwijs.nl/dialoog/ last accessed on 15th June 2020

to a great extent, on its capacity to construct, mediate and disseminate policy ideas, and in the legitimacy that national actors attribute to it in the enactment of such capacity.

Whilst maintaining overall control over the review, interview data reveals ongoing dialogue, ideas-sharing and negotiation between the OECD and MoE teams throughout the process. OECD reviewers were aware of the MoE's strategic commissioning of the review, yet welcomed their 'expert at will image' and appeared eager to help the government explore and recalibrate its system. In this sense, the apparent malleability of the IO is central to its success. At times even, this malleability extended beyond what MoE actors perceived as desirable. Yet, this relationship may be somewhat of a Dutch luxury, resulting from the OECD's respect for the 'good student' – well-aligned with the NPM and SAWA agendas. The fact that a background paper was not commissioned as a preliminary step, and that Dutch citizens were part of the review team also reflects the OECD's confidence in the Netherlands. In countries where this is not the case, the IO might adopt a harder, more patronising approach (see for instance Echavarrri & Peraza 2016). Yet, the adaptability of the RNPE should not be surprising. Autonomy and accountability solutions can be attached to a vast array of problems. Being only loosely coupled (or not coupled) to specific policy issues is a distinct advantage when it comes to promoting a particular policy agenda.

This SAWA agenda is evident throughout the review process and the report itself. It is promoted in key ways; through the use of (pressure-inducing) benchmarking data, through the problematisation of policies considered ill-aligned with SAWA ideals, and through the report's main recommendations. These methods point to the underlying mechanism of multilateral surveillance: decentralisation, standardisation, and test-based accountability have become the 'rules of the game'. In the ways described, RNPEs appear to be an efficient means through which to enforce these rules and transfer normative notions of how educational systems should be organised and governed.

As suggested by Marcussen (2004), the impact of RNPE's on domestic policy can be difficult to determine, and reference to OECD documents and ideas does not necessarily mean that learning, or idea transfer has taken place. Unlike other research, which has focused on contexts where such reviews have had considerable impact – Norway (Steiner-Khamsi et al 2020) and Chile (Bravo 2011, Parcerisa & Falabella 2017) for example – direct policy impact of the 2016 Dutch review appears limited. Interestingly, according to respondents, the one area in which some policy change *was* reported to have occurred following the report, was the area that proved the most contentious during the review process – early childhood education. During the draft feedback stage, issues that were highlighted here seemed to come as a surprise to the MoE, and were seen as overly critical and suggestive of insufficient quality care provision. This possibly indicates that the influence of moral pressure (Drezner 2000), even 'shaming' (Marcussen 2004) is significant. We must too consider that the review was reported to not have delivered the fresh insights the MoE had hoped it would. As previous research has shown, if not

interested in particular findings, the government might act to limit a report's dissemination (Verger et al 2019). Whether or not this happened in the Netherlands is difficult to determine, yet one OECD member did express surprise at the fact that other important reports and significant policy news were also released on RNPE launch day.

A lack of immediate policy reform however does not mean that the review lacked its uses. Findings show the RNPE was used in several ways by the MoE, including for retrospective policy justification, and reframing and redirecting 'undesirable' sentiments. Given time, its use may also be seen in the re-emerging discussions on standardised testing, and the early student streaming system. It is worth noting that not all governments may be able to use RNPE to their advantage in this way. As a well-performing system, with strong institutions and a history of conducting its own rigorous research, the Dutch MoE has the ability to make strategic and selective use of OECD recommendations. An educational system in a less developed economy, or an educational system perceived to be in crisis, may not have this privilege.

To conclude, the study shows that the OECD's capacity to influence national policy depends on indirect and diffuse forms of institutional power that rely on expert knowledge as well as on agenda- and norm-setting mechanisms. Through these mechanisms, whose enactment strongly relies on the socialisation and networking between policy elites and experts, the OECD establishes the limits of educational debates and the types of problems that need to and can be addressed. Member countries are not passive receptors of this agenda. Countries, especially those that are better situated in the OECD system and whose policy perspective is aligned to that of the IO, actively contribute to co-construct the OECD agenda and to consolidate its international legitimacy. Policy reviews are, as has been shown, an empirical manifestation of this form of global governance in the making.

REFERENCES

Bieber, T., & Martens, K. "The OECD PISA study as a soft power in education? Lessons from Switzerland and the US." *European Journal of Education* 46.1 (2011): 101-116.

Breakspear, S. (2012). *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance.*

Broome, A., & Quirk, J. (2015). The politics of numbers: the normative agendas of global benchmarking. Review of International Studies, 41, pp 813-818

Browes, N., & K Altinyelken, H. (2019). The instrumentation of test-based accountability in the autonomous Dutch system. *Journal of Education Policy*, 1-22.

Cox, R. & Jacobson, H. (1973). *The Anatomy of Influence: Decision Making in International Organization* (New Haven: Yale University Press)

Drezner, D. W. (2000), ‘Bargaining, enforcement, and sanctions: when is cooperation counterproductive?’, International Organization54(1), 73–102.

Echávarri, J., & Peraza, C. (2017). Modernizing Schools in Mexico: The Rise of Teacher Assessment and School-Based Management Policies. *education policy analysis archives*, *25*(90).

Emons, W., Glas, C., Berding-Oldersma, P. (2016). *Rapportage Vergelijkbaarheid Eindtoetsen.* Available at:https://www.eerstekamer.nl/overig/20170126/rapportage_vergelijkbaarheid/f=y.pdf

Fischman, G. E., Topper, A. M., Silova, I., Goebel, J., & Holloway, J. L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of education policy*, *34*(4), 470-499.

Glenn, C., L. & De Groof, J. (2005). *Balancing freedom, autonomy, and accountability in education.* Nijmegen, Wolf Legal Publishers

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

Grek, S. (2014). OECD as a site of coproduction: European education governance and the new politics of ‘policy mobilization’. *Critical Policy Studies*, *8*(3), 266-281.

Grek, S. (2016). Beyond the standardisation vs. contextualisation debate: The role of the OECD in European education governance. Paper presented at the International Organizations and the Globalization of Public Instruments and Ideas (IOGLOB) Seminar, Strasbourg.

Grek, S. (2017). Socialisation, learning and the OECD’s Reviews of National Policies for Education: the case of Sweden, Critical Studies in Education, 58:3, 295-310

Halton, M. J. (2003). Benchmarking: Another attempt to introduce market-oriented policies into Irish second-level education? Pedagogy, Culture and Society, 11(3), 331–351.

Inspectorate of Education, Ministry of Education, Culture & Science (2017). *The state of education 2015/16* (Eng.) www.destaatvanhetonderwijs.nl.

Istance, D. (2011). Education at OECD: recent themes and recommendations. European Journal of Education, 46(1), 87–100.

Komatsu, H., & Rappleye, J. (2019). Refuting the OECD-World Bank development narrative: was East Asia’s ‘Economic Miracle’ primarily driven by education quality and cognitive skills?. *Globalisation, Societies and Education*, *17*(2), 101-116.

Lewis, S. Sellar, S. & Lingard, B. (2016) "PISA for Schools: Topological Rationality and New Spaces of the OECD’s Global Educational Governance," Comparative Education Review 60, no. 1: 27-57.

Mahon, R., & McBride, S. (2009). Introduction. In Mahon, R., & McBride, S. (Eds.) *The OECD and transnational governance*. UBC Press. 3-24.

Marcussen, M. (2004). Chapter 2, Multilateral surveillance and the OECD: playing the idea game. *The OECD and European welfare states*. Armingeon, Klaus, and Michelle Beyeler, eds. Edward Elgar Publishing

Martell, L. (2007). The third wave in globalization theory. *International studies review*, *9*(2), 173-196.

Ministerie van Onderwijs, Cultuur en Wetenschap ‘MoE’. (2016). *Kamerbrief over resultaten doorlichting Nederlands onderwijsstelsel*. Reference Number: 939852.

https://www.rijksoverheid.nl/onderwerpen/financiering-onderwijs/documenten/kamerstukken/2016/05/25/kamerbrief-over-resultaten-doorlichting-nederlands-onderwijsstelsel

Møller, J., & Skedsmo, G. (2013). Modernising education: New Public Management reform in the Norwegian education system. Journal of Educational Administration and History, 45(4), 336–353.

Mundy, K., Green, A., Lingard, B., & Verger, A. (Eds.). (2016). *Handbook of global education policy*. London: John Wiley & Sons.

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

Nusche, D Braun, H., Gábor, H., & Santiago, P. (2014). OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing. http://dx.doi.org/10.1787/9789264211940-en

OECD. (2011). PISA in focus, no. 9. School autonomy and accountability: Are they related to student performance? Paris: OECD Publishing. https://doi.org/10.1787/5k9h362kcx9w-en.

OECD (2014). Education Policy Outlook, Netherlands http://www.oecd.org/education/EDUCATION%20POLICY%20OUTLOOK_NETHERLANDS_EN%20.pdf

OECD (2016). Netherlands 2016: Foundations for the Future, Reviews of National Policies for Education, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264257658-en

Onderwijsraad (2010).Vroeg of laat Advies over de vroege selectie in het Nederlandse onderwijs, Nr. 20100040/969, Den Haag.

Pagani, F. (2002). Peer review as a tool for co-operation and change: An analysis of an OECD working method. African Security Review, 11(4), 15–24.

Pal, L. (2012). *Frontiers of governance: The OECD and global public management reform*. Springer.

Parcerisa, L., & Falabella, A. (2017). The consolidation of the evaluative state through accountability policies: Trajectory, enactment and tensions in the Chilean education system. *Education Policy Analysis Archives*, 25(89), 1–27.

Porter, T., & Webb, M. (2008). The role of the OECD in the orchestration of global knowledge networks. *The OECD and transnational governance*, 43-59.

Sellar, S., & Lingard, B. (2013). The OECD and global governance in education. *Journal of Education Policy*, 28(5), 710–725.

Schuller, T. (2005). Constructing international policy research: The role of CERI/OECD. European Educational Research Journal, 4(3), 170–180.

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

Verger, A., Fontdevila, C. and Parcerisa, L. (2019). Constructing school autonomy with accountability as a global policy model: A focus on OECD governance mechanisms in Ydesen, C. (ed). The OECD's Historical Rise in Education: The Formation of a Global Governing Complex. London: Palgrave Macmillan. pp. 219-243.

Waldow, F., & Steiner-Khamsi, G. (Eds.). (2019). *Understanding PISA's attractiveness: Critical analyses in comparative policy studies*. Bloomsbury Publishing.

Waslander, S., Hooge, E., & Drewes, T. (2016). Steering Dynamics in the Dutch Education System. *European Journal of Education 51*(4), 478-494.

Wöbmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). School accountability, autonomy, choice, and the level of student achievement.

CHAPTER 4

POLICY ENACTMENT I:

Professionalism in the Era of Accountability: Role Discrepancy & Responses amongst Teachers in the Netherlands

This chapter is a pre-print version of: Natalie Browes & Hülya K Altinyelken (forth). Professionalism in the Era of Accountability: Role Discrepancy & Responses amongst Teachers in the Netherlands. Submitted to British Journal of Sociology on 29th July 2020

INTRODUCTION

The neoliberal agenda has acted as a major catalyst of change across the professions which have seen a growing adoption of business values in the quest for efficiency and effectiveness (Ball 2016; Muzio, Brock, & Suddaby 2013). This has not only transformed the requirements and roles of professionals, but has also transformed how they see themselves and the nature and purpose of their work. This is no truer than in the field of education, where managerial reforms have increased decisions taken at the school level, while external 'surveillance' mechanisms ensure that stakeholders remain accountable (Verger, Fontdevila & Parcerisa 2019). Primarily, this accountability hinges on the attainment of core standards, measured by way of performance 'outcomes'. Certainly at a policy level, rather than act as a proxy of quality, test-based performance indicators have therefore come to represent quality. As well as being expected to 'perform' (see Ball 2003), teachers are also expected to account for their work more broadly through the ritual (and predominately recorded), planning, monitoring, and evaluation of teaching and learning. These policy tools are collectively referred to throughout the paper as 'performance-based accountability' or 'PBA.'

These reforms have fundamentally changed the concept of quality education and what it means to be a teaching professional. 'Successful education systems' are those topping the international league tables in large-scale students assessments; 'successful schools' are those that outperform their neighbours in standardised tests; and 'successful teachers' are those that add the most 'value' (grade points) to their

students. A considerable amount of research has examined the transformation of the professions over the decades (see for example: Evetts' 2003; Faulconbridge & Muzio 2012) and teachers' changing roles and identities through their enactment of these reforms constitutes a significant part of this (Day 2002; Sachs 2001; Valli & Buese 2007). The degree to which teachers' beliefs and practices align with policy varies (Ball 2003; Holloway & Brass 2018; Moore, Edwards, Haplin & George 2002), linked to a complex interplay of factors at various levels. What is clear is that, like all professionals, teachers are not passive implementers of policy: they are key actors who shape policy through processes of interpretation and translation (Ball, Maguire, Braun & Hoskins 2011). The policy demands on teachers may be at odds with their own understanding of what it means to be a good teacher, inevitably resulting in compromises being made – either on behalf of the teacher's professional beliefs, on behalf of the policy, or both.

Based on 20 interviews in six primary schools in the Netherlands, this study examines teachers' professionalism in the era of (performance-based) accountability, using the concept of *role discrepancy*. More specifically, it seeks to discover whether teachers experience a divergence between what they do – those tasks they actually engage in (or are expected to engage in) at work, and what they believe – those tasks they consider at the heart of good teaching. It is framed around three questions:

I. What are teachers' experiences of their work tasks: what tasks do they engage in, to what extent are they 'performative,' and how do they feel about these tasks?
II. Do teachers experience role discrepancy and if so, what work-management approaches are adopted?
III. What does the examination of teachers' roles, practices, and beliefs, reveal about their professional identities in the era of PBA?

The study forms part of a broader research project [name] which examines globally-spread school autonomy and accountability policies and their development and enactment in various countries [website]. The paper aims to contribute to our understanding of the enactment of accountability reforms, and more broadly, to the possible impact of PBA on the teaching profession, in a number of ways. Regarding the study's *conceptual contributions,* although the core aspects of 'beliefs' and 'practices' cut to the heart of professionalism (Demirkasimoğlu 2010; Hendrikx 2019), the concept of 'role discrepancy' is still rather limited. While it has been applied (quantitatively) to some professions, including nursing (Takase, Maude & Manias 2006a & 2006b) and school social workers (Agresta 2006), its use in teaching, and as part of a qualitative exploratory study is largely absent. Further, the paper contributes new conceptualisations of policy enactment that go beyond the dichotomous labels of 'resistance' and 'compliance,' providing a deeper understanding of teachers' working realities. Regarding the study's *contextual contributions,* while widespread in certain neoliberal, high-stakes contexts (particularly England and the US), there is insufficient research examining the enactment of

PBA in systems with different institutional traditions and differently-constructed policies, and therefore insufficient understanding of which formulations of PBA produce which effects. Further, in the Netherlands, little research has focused on the enactment of such policies and their impact on primary school teachers (for research at the secondary level, see Hendrikx, 2019), and yet, the Dutch policy changes imposed over the last dozen years have fundamentally reshaped the primary education system (Browes & Altinyelken 2019). Finally, the study has important *policy and social implications.* Research has found that a change in teacher's roles and excessive role regulation can result in: 'job dissatisfaction, reduced commitment, burnout, loss of self- esteem, and early departure from the profession' (Calderhead 2001 in Valli & Buese 2007 p.521). Similarly, role discrepancy has been found to be linked to intention to quit (Takase 2006b) and job dissatisfaction (Agresta 2006). Given the instabilities and uncertainties currently surrounding the teaching profession in the Netherlands (including ongoing national strikes and crippling teacher shortages), a better understanding of teachers' practices and beliefs in the current policy environment is crucial.

PROFESSIONALISM, ROLE DISCREPENCY & ENACTMENT

Professionalism is an elusive and changing concept. It is multi-faceted, multi-scalar and multi-purpose. According to Evetts (2003; 2009) the term can be understood as a set of normative values, practices and discourses around a given profession which can generate and facilitate alterations of the occupation. While efforts to exclusively define 'the professions' are somewhat outdated (Evetts 2003; Muzio et al 2013), certain key aspects have traditionally been associated with this group of workers including; autonomy and the ability to make discretionary judgements, a 'public service' dimension, and (development of) competency and 'expertise' (Demirkasimoğlu 2010), as well as collegiality and collaboration (Clement and Vandenberghe 2000). Evetts (2009), highlighting the dynamic nature of professionalism, has identified two 'types'; *occupational professionalism*:- understood as produced from within and defined by collegiality, trust and autonomy, and *organizational professionalism*:- produced through an external discourse of control and defined by managerialism, standardisation, and external accountability.

A number of studies have examined the concept of teacher professionalism and, more specifically, its metamorphosis through educational reform (Day 2002; Hargreaves 2000). Essentially, we can understand professionalism here as the development of the individual teacher in order to improve quality and standards of practice for the student, not to be confused with *professionalization*, where the concern is with (improving) teachers' status, standing, regard and reward (Hargreaves 2000). Based on the Anglo-American experience, Hargreaves (2000) describes professionalism in teaching as passing through several stages; from the pre-professional, to the autonomous professional, the collegial

professional, and finally, to the post-professional of today, where teachers struggle to counter centralized curricula, testing regimes, external surveillance, and the economic imperatives of marketization.

Whether 'post-professionalism' (Ball 2003; Hargreaves 2000), 'organizational-professionalism' (Evetts 2009), or one of several similar constructs, there is a strong argument that professionalism today is shaped by new public management values; efficiency, standardisation, and competition. This is challenging teachers' broader identities: 'teachers and other public services workers succeed only by satisfying and complying with others' definitions of their work" (Day 2002 p.682), yet teachers' own perceptions of their work might diverge from these definitions, resulting in a mismatch between (required) practices and beliefs. The distinction between what teachers do and what they feel they should do (i.e. their professional values), can be understood as *role discrepancy*: 'the incongruence between their [the professionals'] ideal roles and the roles they actually engage in at work' (Takase 2006a).

This concept has been applied to a number of professions, including nursing (Takase, Maude & Manias 2006a & 2006b), school social workers (Agresta 2006), and occupational therapists (Lloyd, King & Mckenna 2004). These studies have found role discrepancy to be prevalent. More specifically, Agresta (2006) notes that school social workers wanted to spend less time on routine tasks, particularly report writing, and more time on those tasks considered at the core of their profession: individual and group counselling. Lloyd et al (2004) similarly found that occupational therapists desired to spend more time on their areas of specialism. These quantitative studies have also shown experience of role discrepancy to be negatively associated with job satisfaction (Agresta 2006) and positively associated with intention to quit (Takase et el 2006b).

Application of the role discrepancy concept to the teaching profession has not been found, at least not within major English literature databases, yet research has built on similar ideas. Studies, particularly US-based, have revealed the largely negative impacts of high-stakes PBA on teachers' roles. Olivant (2015) saw teachers struggling with time pressures, diminished autonomy and professionalism, and creativity in the classroom. Valli and Buese (2007) found teachers who were governed by prescribed curricula and standardised testing to be experiencing role increase, role intensification, and role expansion, with negative impacts on pedagogy, 'professional well-being' and student relationships. The authors' discussion of 'role expectations' and the conflicts surrounding this, are particularly close to the notion of role discrepancy.

Teacher response to PBA has also received considerable attention, with studies showing significant variation in enactment strategies. Achinstein and Ogawa (2006) examined the 'principled resistance' of teachers in a Californian policy context, who refused to follow a heavily-prescriptive literacy programme on ideological grounds. Other, more subtle responses, include Perryman, Ball, Maguire & Braun's (2011) notion of 'policy evasion', whereby, rather than outright resistance, teachers in English

secondary schools were constrained to selecting the aspects of policy considered (not) worth implementing. Similarly, Moore et al (2002), also observing the English context, found teachers – whether adopting more 'compliant' or 'resistant' positions – to adopt an eclectic, pragmatic approach to policy enactment. The authors use the term 'contingent pragmatism' to describe teachers who see their practices as somewhat of an enforced survival strategy, and 'principled pragmatism' for those who better reconcile their beliefs and practices, often justified through their students' performance outcomes. In the Dutch context, a recent study in secondary schools (Hendrikx 2019) found that mismatches between teachers' beliefs and practices were common, resulting in compromises being made on one side or the other. The (qualitative) study also identified a third response, in which teachers reduced their contracted work hours without reducing their actual work hours, enabling them to engage in the work meaningful to them whilst still fulfilling obligations.

Similar studies in the enactment field have indicated that, as well as factors at the school, locale, and system levels, at the individual level a teacher's professional experience might play an important role in their perception and experience of daily work tasks. It has been found that more experienced teachers feel more frustrated with their roles and experience of PBA, perhaps as their training and formative teaching years were shaped in a different policy environment (Day & Smethem 2009; Holloway & Brass 2018) or perhaps because they are simply worn down by 'repetitive educational change' (Hargreaves 2005). Teachers who entered the profession more recently (for whom, such demands have been a constant reality), were found to have had their professional beliefs and identities inherently shaped by this policy environment (Holloway & Brass 2018; Wilkins 2011).

Notably, Takase et al's study (2006a), which explicitly tested the relationship between role discrepancy and years of (clinical) experience, presents a contrasting premise:

> … it has been assumed that experienced nurses perceive less role discrepancy than inexperienced nurses, either because the former adjust themselves to their actual practice or because they have the expertise to improve their practice… (p.751)

In fact, the study's findings revealed that more- and less-experienced nurses experienced similar levels of role discrepancy, yet more-experienced nurses reported more positively on their work roles.

TEACHERS & ACCOUNTABILITY IN THE NETHERLANDS

The Dutch Teaching Profession

Over the years, the teaching profession in the Netherlands has been steered in different directions. It has recently been experiencing severe teacher shortages, particularly in urban areas, and ongoing teacher strikes over high workload and insufficient pay (Inspectorate of Education 2019). With a perception that

teaching is no longer the respected profession it once was, attracting new recruits is proving a difficult task (only 10 per cent of Dutch teachers are under the age of 30) (OECD 2019). According to recent TALIS secondary education data (OECD 2018), teaching was a first career choice for only 53 per cent of Dutch teachers, considerably lower than the 67 per cent TALIS average. While teachers' salaries are somewhat higher than average for OECD countries, they are lower than those of other Dutch professionals (OECD 2019), particularly at the primary school level.

Over the last decade, the government has been making a clear attempt to professionalise teaching; to raise its status and desirability and to encourage into the profession, more men and candidates with academic backgrounds. Initiatives include; reduced training time for those with an academic or professional background, 'the hybrid teacher' (encouraging professionals from other fields to combine their work with teaching), reduced tuition fees for teacher training, grants for teachers to complete Masters or PhDs, a teachers' 'competency framework' (a professional statute and standard for teachers), and the 'teachers' register' whereby teachers must complete a minimum amount of (certified) professional development hours (see 'de lerarenagenda 2013-2020'). In response to the aforementioned strikes, primary teachers have recently been granted a pay rise along with extra money at the school level to help reduce teacher workload. Teachers considered these steps insufficient and at the time of writing, intermittent strikes continue.

The goal of these initiatives is teacher professionalization (concerned with status), not professionalism (concerned with practice) (Hoyle 2001). One reason for this may be the high levels of decentralisation and autonomy in the Dutch system, inhibiting government involvement in the teaching *process.* Indeed at the input level, legislation is scarce, including a lack of a national curriculum and a centralised system of teacher appraisal. While teaching quality forms an important part of the Inspectorate's framework, this is only evaluated at the school level. Yet, the (governance of the) Dutch system has been described as one of the most complex in the world (Waslander, Hooge & Drewes 2016) and despite this 'school autonomy' label, surveillance operates as a chain. As well as intermediary 'sector organisations,' such as the council for primary education, which use a variety of steering mechanisms to manipulate school practice, school-boards (under strict accountability themselves) implement various measures and place pressure on school principals to ensure sufficient school performance. These may be formal requirements – since 2006, boards must keep competency files on teachers and ensure that annual performance reviews take place (Nusche, Braun, Halász & Santiago 2014) – but equally, may be additional, protective measures. These layered structures mean that teachers are operating under internal, as well as external surveillance (see Skerritt 2020).

Accountability & Performativity

Despite little *direct* government involvement in their work, teachers are certainly not free from the demands of performativity. PBA policy tools have been introduced incrementally, most significantly between the period 2008-2014 (Browes & Altinyelken 2019). Core learning standards, measured by

compulsory standardised testing throughout primary education, manipulate teachers' goals, values, and teaching practices. This is amplified by the presence of administrative and market tools that increase the stakes attached to school performance: Average test scores are published and compared to national averages. Parents are encouraged to consult this data, whether choosing a school or 'keeping an eye' on one. Performing below average for three consecutive years, a school will be labelled 'very weak' and subjected to an intensive and extensive inspection process (Inspectie van het Onderwijs 2020). Tests also have significant stakes for students given that they are tracked into secondary education based on their performance at primary level. This is an added pressure for teachers (Browes forth.).

Companies have developed subject packages or 'methods' to support schools in meeting government outcome requirements. These include detailed syllabi, textbooks, and end of unit assessments. It is extremely common for schools to purchase and follow these methods, particularly in the core learning areas. Schools are also required to show data-oriented working and the use of effective quality assurance systems. This has led to the purchasing of student administrative systems, on which teachers are expected to record student progress and performance, as well as other student-related matters such as behaviour and meetings with parents. Also indicative of a data and documentation culture – producing, updating, and evaluating 'class plans' has become a core part of teachers' work. Not to be confused with daily lesson plans, these are longer-term, standards-oriented plans for each school period, often broken down by teachers into weekly plans. Class plans (usually) differentiate students by ability and organise learning accordingly. Ability groupings and lesson content are heavily influenced by standardised tests. Contrary to common belief, class plans are not a government requirement. This fact has been lost under layers of governance, and the plans have come to play a pivotal role in the organisation of learning and often in schools' own evaluation of teachers (Browes forth).

METHODS

This study aims to uncover new ways of understanding teacher enactment of, and beliefs towards, PBA through a qualitative research strategy. For reasons outlined in the introduction, teachers in Dutch primary schools were sampled to explore the impact of 'global' education trends through a case-study approach (Yin 1984). Subscribing to the work of authors such as Stephen Ball, Anette Braun & Meg Maguire, enactment is understood as context specific and teachers are seen as pivotal actors in the interpretation and translation of policy. For this reason, a realist evaluation approach was adopted to guide data collection and analysis (see Pawson & Tilley 1997), acknowledging the, often unobservable, mechanisms that impact teachers' (different) policy experiences.

Interviews were conducted with 20 teachers in six primary schools across three small cities in the central west 'Randstad' region.[23] Three of the schools are public 'openbaar,' two independent 'bijzonder' and one, a merger school, is described as both. All are fully government-funded, yet the public schools are also government-established meaning they do not align with any specific philosophical, religious, or social movement (also true of the merger school). The remaining two independent schools are 'Protestant', yet background discussions with principals revealed this to be more reflective of the history of Dutch education (Strum, Groenendijk, Kruithof & Rens 1998) than of school culture. The schools varied in terms of their student populations and their approaches to education. Two of the schools had recently adopted a project-based, student-centred approach to learning, in which teachers took on a 'coaching' role. A third school had started using this approach in some non-core subjects. Three of the schools are located in affluent areas with homogenous 'native Dutch' populations, two in 'mixed' neighbourhoods, and one in a majority immigrant area.

Due to teacher shortages and reportedly high workloads, securing participation was challenging. Four of the schools were contacted 'cold' and the remaining two through the network of the second author. Participant selection prioritized teachers from the upper level[24] – groups 6, 7, & 8 – where PBA is at its most pronounced (stakes attached to testing are higher for students and for the school than in previous years). One respondent had recently dropped her teaching responsibilities to focus on a management position, and is therefore cited as a 'vice-principal'. In total, the sample consists of 13 female and 7 male teachers. Years of teaching experience varied considerably, from less than one year to over 25 years. Four teachers had less than five years' experience, and five teachers over 15 years'. For an overview of research participants and the schools in which they work, see table 3.

Interviews were semi-structured and averaged 40 minutes. They were transcribed verbatim and coded using software for qualitative data analysis. All participants have been given pseudonyms to preserve their anonymity. The interview script was guided by the study's research questions and realist evaluation perspective, and by a backlog of previous work which has examined the mechanisms and impacts of accountability and managerialism (for example, Evetts 2009; Verger & Parcerisa 2017). Themes included; 'work schedules & work tasks,' 'accountability,' 'ideal work tasks & working strategies,' 'professional image,' and 'school environment.' It was semi-structured, allowing room for probing and ad hoc questions to adapt to teacher's individual situations and experiences. Coding was initially guided by the interview script and research questions, yet, it was also iterative in the sense that interviews were (re-)read to uncover themes and reflect the meaning behind teachers' responses, with the coding protocol

---

[23] An area of high urbanity encapsulating the country's four biggest cities.

[24] It is common however for teachers to rotate groups every year or few years (usually keeping within the lower, middle, or upper level).

revised accordingly. The uncovering of various types of 'working approaches' in response to PBA for example, initiated a new round of coding.

Table 3: Participating Teachers & their Schools

School No.	Province	Type of school	No. interviews	Name* & position	Years Teaching
1	North Holland	Public	4	Sanne, group 3	12
				Emma, group 6	10
				Lotte, group 8	4
				Roel, group 7/8	<1
2	North Holland	Public	4	Mirjam, group 8	12
				Robbie, group 5	4
				Julia, group 6	5
				Emily, group 7/8	17
3	North Holland	Public	3	Lisa, group 2	10
				Thom, group 7	12
				Nick, group 6	11
4	North Holland	Public / Independent: General	4	Sophia, group 8	17
				Matthijs, group 5	14
				Femke, group 7	18
				Daan, group 8	19
5	Flevoland	Independent: Protestant	3	Tessa, group 7	7
				Lynn, group 8	14
				Martine, VP (previously group 6)	6
6	South Holland	Independent: Protestant	2	Lucas, group 7	1
				Lianne, group 8	25
Total:			**20**		

FINDINGS

Findings are organised around the three research questions. The first section examines teachers' experiences of their work tasks and to what extent these tasks are performative. As well as teacher practices (how they spend their time) it also explores teacher beliefs (how they would ideally spend their time). The second focuses on the differences between these beliefs and practices, asking whether role discrepancy is experienced and if so, how this impacts teachers' working approaches. Finally, section three considers, in this PBA environment, how teachers might be (re)orienting their sense of professionalism.

Teachers' Tasks: Practices & Beliefs

The majority of teachers' in-school time is spent teaching their students. Teachers across the six schools follow similar routines; dedicating (at least) morning lessons to the core subjects in an effort to better utilise student concentration. Respondents unanimously experience teaching time as intense: following crowded teaching plans, prescribed curricula, and fixed attainment goals. For some, this resulted in a sense that there is no time for flexibility, fun, and exploration in the classroom.

> In these developments in the last 10 years, we've become more and more slaves of the method; 'we need to do lesson 5 and then you need to go to this topic and then… .' There's very little room to do something for fun, to do something because it is just interesting but it doesn't add up to any measurable results.
>
> (Emma, group 6, school 1)

Several teachers reported that these demanding schedules force a choice; between 'keeping up' and student understanding, and results in constant 'clock watching'. Within these teaching schedules, the assessed areas of learning appear to dominate. Many respondents desired to spend more time teaching topics and subjects not part of the core standards. Teachers in four out of the six schools expressed concern that creative subjects such as music and art are being squeezed out. Teachers also want the time and flexibility to be able to address ad hoc socio-cultural or behavioural issues, such as bullying. Yet here, the data reveals a perceptual difference, seemingly stemming from the value that teachers attribute to standards-based learning: whereas respondents generally agreed that more time should be spent on creative subjects, some thought this should come at the expense of core subjects, others, who see these subjects as critical part of their students' education, did not.

> But of course we sometimes say, '*but how? when?'* If I take an hour and a half painting that means I haven't done dictation, so next week I have to squeeze it in. And for some kids it's okay to miss a little lesson, but for the weaker kids, they need that extra time, we can't say 'okay, we're all going to paint, but *you* can't! - *You* have to sit and do your extra practice.'
>
> (Sanne, group 3, school 1)

Teachers working in schools which adopt project-based approaches generally spoke more favourably of their teaching experiences. Martine, who had recently moved schools, had seen the benefit of this approach first hand:

> [At my previous school] the teacher was leading, and not the children. There, I had a tight schedule; math, spelling, language. And the children there, they did not get the space to discover and to explore.
>
> (Martine, vice-principal, school 5)

It was common for teachers in the more 'pedagogically-traditional' to perceive their classes as regimented and inflexible. These teachers felt their students would benefit from a more student-centred approach and many desired a move to project-based learning in particular.

In four of the participating schools, the teaching period has become intensified over recent years due to a shortened school day: the result of devolved responsibilities that have given school boards the flexibility to choose opening hours. While this shortened day is preferred by some, others were not so keen. Robbie for example, believed these condensed teaching hours were contributing to an erosion of collegiality and ultimately, an increased workload:

> They [the students] are off at 2 o'clock, but what about finishing at 3 and having an extra 45 minutes of break? (...) And it would save me time, as it would give me time to talk to colleagues, to be together and talk about ideas (…) 'I have a measurement lesson in maths, what can I do – any tips?' So that would be good, and it would mean more bonding and you exchange ideas more to reduce the workload. I think that would be my ideal thing.
>
> (Robbie, group 5, school 2)

Other teaching-related tasks are scheduled for afternoons, once students have gone home. In particular, this includes lesson planning for the following day and checking and marking students' home and class work. The majority of respondents felt they have insufficient time for these tasks due to other responsibilities; meetings, responding to parents' emails, and administrative duties. Most of these respondents desired more time for lesson-planning in particular, believing this would enable their students to benefit from more impactful classes and result in a more enjoyable teaching experience.

Beyond these core teaching tasks, test-based activities occupy a considerable portion of teachers' schedules and the vast majority of respondents experience a workload increase during testing periods. Test-based tasks chiefly consist of; preparing for tests (readying students, organising papers, and preparing the classroom), inputting and analysing test data (the majority of tests are done by hand and must subsequently be digitised) and 'acting on results' (meetings with management, group meetings with colleagues, parent meetings, student reports, ability grouping and teaching content and timetables are all based around results). Testing is generally seen as a necessary part of school life, and although numerous respondents felt its role has become too dominant, it is still understood as a useful way to establish goals, assess progress and tailor teaching. Testing itself does not clash with teachers' professional values, yet for some, the *centrality* of testing, and the administrative burdens connected to it, do.

These administrative or more specifically 'documentation' tasks reportedly demand a considerable amount of teachers' time. Their general purpose appears to be teacher surveillance and accountability. Teachers are required to show plans and records of students' (differentiated) work, and regularly evaluate and modify these plans to demonstrate they are aware, data-oriented, and responsive. Opinions about the value of such tasks vary. A small minority of respondents had a positive outlook, reporting them to be a necessary way to stay 'on track' and abreast of their students' progress. A much larger number considered these documentation demands excessive and never-ending. Some teachers in this second group found the tasks entirely unnecessary:

> Well, the principal of the school wants that [daily report]. I asked why he thinks that's necessary, I said 'It's total nonsense! It's absolutely not necessary, because I'm the only teacher here in my class – I work 5 days.' I can imagine that if you have another colleague – you work 3 days and they

> work 2 – you have to show what you've done and explain the difficulties, but I'm by myself, so should I get into a dialogue with myself? That's nonsense. *I know*, I have in *my head* what's happened.
>
> (Matthijs, group 5, school 4)

Finally, teachers' communication with parents was commonly reported to be highly-demanding and a generally undesirable task. Notions parents may have of the 'accessible teacher' are also symptomatic of the PBA agenda, connected to the goals of transparency and participative accountability. Respondents felt, at times, overwhelmed by parents' questions and demands: for less experienced teachers, this was something they had not expected. For those in the profession longer, it was something they had seen change.

> Twenty years ago it was easy. When I started, internet was just upcoming, e-mailing was not a priority. But now parents don't come to your class anymore, [they] just send an e-mail. Sometimes, an angry e-mail. You feel that the tone is angry, but you never know in what mood they texted, always guessing. And that is, that irritates me the most nowadays. Just sending problems.
>
> (Daan, group 8, school 4)

Role Discrepancy & Teacher Response

The account of teachers' experiences provided above, reveals strong and recurrent themes of heavy workloads, task-management pressures, and, for many, insufficient time in the working week. Essentially, teachers appear to be experiencing a task-overload. These tasks are heavily influenced by the PBA agenda: teaching, testing, and administrative demands are ultimately shaped by the pursuit of good, or at least sufficient, test scores.

A small minority of teachers (3 out of the 20 respondents), accepted the importance of this agenda, to the extent that they expressed an *alignment* between the tasks they engage with at work and the tasks they considered at the core of good teaching: essentially, these teachers have seemingly aligned their practises and beliefs: This group of respondents attached importance to PBA tools: Learning standards were considered to keep them focused and abreast of what is important, standardised tests were perceived as an essential way of checking their students' and their own progress, and record-keeping and careful planning were considered to make them more effective teachers, better able to keep track of their students and foster improvement.

> Well in your lessons of course you try to see if everybody gets the [learning] goals, and for me it's to see, *what are* the goals? And you need to put those in a report for home [parents] and it's like… well it keeps you *alert.*
>
> (Lotte, group 8, school 1)

While the numbers are small, it is still noteworthy that all three of these teachers were in the very early stages of their careers (teaching for less than five years). They were not blind to the issues that an over-focus on performance could generate and still would have liked more time for particular tasks, yet they wanted their students to achieve to the best of their abilities – believing this to be the essence of teaching – and saw the contribution that performative tasks made to this goal. In this way, their students' test scores were considered to reflect on their own professionalism.

The remaining majority of respondents can be understood to be experiencing a mismatch between those tasks they are (expected to) engage with, and those tasks they considered at the core of good teaching. Performative tasks were perceived to be numerous and highly demanding. Respondents felt these tasks detracted from the those they considered to be at the heart of the profession; planning and delivering engaging, thought-provoking lessons, and helping students to maximise their abilities. Teachers discussed various responses to these role discrepancies. Some, position role discrepancy as something that can be 'managed;' with teachers finding time for performative tasks *as well as* time for those tasks they felt to be most important. Others, position it as something to submit to; with teachers feeling forced to choose between tasks. Attached to each response are particular working attitudes and approaches. While most respondents can be understood as adopting one position or the other (being fairly evenly spread between the two), and taking one particular approach to their work, a minority of teachers can better be understood as moving between the two.

Role discrepancy 'management'

Some teachers described trying to manage role discrepancy by adapting their way of working to make space for their beliefs as well as required practices. In this way, all tasks are integrated into teachers' working schedules. Respondents spoke of one of two ways in which this was achieved.

The first way is through the *integration* of tasks. Despite not attaching particular importance to the performative tasks required of them, some respondents felt able to integrate these tasks into their work routines in such a way that they were not perceived to impinge on tasks considered more important. Four teaches spoke of their work approach in this way. They seem to have adapted fairly smoothly to the 'new realities' of teaching by accepting documentation and data-based tasks as part of the job, without feeling compelled to make them a priority, or to make (significant) compromises.

> Well there are things that *have to* be done – things with parents have to be done, tests have to be checked, but for example, a meeting doesn't have to be prepared. So that's the reason, that's the difference, and most of the time, perhaps it's a bit braggy, but I have very good memory so everything that has been said I can recall weeks later, so when I have conversations with parents, I don't have to put them in the computer immediately, I can do it a few weeks later (…).
>
> (Lucas, group 7, school 6)

While at times expressing frustration with managerial demands, these respondents did not see them as worthy of real concern. Teachers who employed this approach had a range of teaching experience; from less than five years to almost 20. Interestingly however, all were male and held positions of responsibility in ICT. Indeed, a technological confidence and an-ease-of-use of the digital systems that have become part of teachers' daily work life, might be central to this more relaxed, integrative approach to administrative performative demands.

Second, were respondents who spoke of task *accumulation*; being able to meaningfully engage with both lesser-valued but required tasks, and the tasks most important to them by incorporating out-of-hours work as a fixed part of their weekly schedules. Most often, tasks undertaken out-of-hours were core teaching tasks (lesson planning and checking and marking students' work) that teachers reported to have insufficient time for during school hours.

> … because we really like to *teach* children something and when we can manage that, and we can see the *fun* of it and *see* it with the kids, well, that's so nice. So that's [lesson planning] something I do on the weekends and that's why I do it on the weekend, because I really like teaching.
>
> (Sophia, group 8, school 4)

Role discrepancy 'capitulation'

A significant number of respondents (between one third and a half) felt, at times, forced into a position of compromise: having to prioritise tasks and choose between (required) practices and those they considered central to good teaching. This was particularly true of those teachers who had made a conscious decision not to work regularly out-of-hours (most of whom had suffered previously from burnout). In these situations, two contrasting working approaches were reported. Firstly, those teachers who reported to economise on required performative tasks in order to make time for other tasks; tasks they considered central to student development. This can be understood as a form of policy *evasion*.

> *Interviewer:* So which tasks do you tend to prioritise then?
>
> *Julia:* Everything that's most important for the children. […] So *preparing* my lesson I think is more important than *typing down* the plan for the whole period.
>
> (Julia, group 6, school 2)

The term 'evasion' is borrowed from Perryman et al (2011). This response appears to stem as much, if not more, from practicality as it does ideology: as reported by Lisa, cutting 'non-core' tasks is sometimes the only option.

> The middle testing period was in January, and then [again] in May or June. And halfway through you have to make an evaluation of your plans… [whispering] and I didn't do that. So I said to [names of management], 'I have no time for it – I have *a lot* of tasks *next to* teaching, I have only two days with an extra assistant, so when do I do it?' And it is in my head… it must be enough!
>
> (Lisa, group 2, school 3)

The ability to adopt this approach appears largely dependent on school management. Certain administrative tasks could be disregarded only because teachers had the trust and support from their school management to do so.

> But I think if I would work at a different school, that would be different. Then you really have to put so much more time in it. But that is something that I would like to do even less of… stupid plans and those kind of things… I never look at those plans when I am teaching.
>
> (Nick, group 6, school 3)

Secondly, teachers who felt consistently forced to economise on tasks they value and consider central to good teaching, expressed a strong sense of *resignation*. These respondents reported a frustration with their everyday work; drowning in lists of daily demands and a sense of constraint; forced to compromise on what and how to teach. Martine (now vice-principal) cited this as her reason for leaving her previous teaching job.

> For a really long time, I tried to do both, but, it was impossible. So you would go with what the management wanted.
>
> (Martine, vice-principal, school 5)

More specifically, this 'response' was characterised by a strong feeling from teachers that demanding workloads have resulted in insufficient time to plan the sort of lessons they want, and a lack of opportunity to teach the lessons they consider really stimulating and beneficial to students. It was primarily reported by more experienced teachers, who spoke of the changes they had seen in the profession and the growing demands upon them. Matthijs, teaching for 14 years, explained that this was not what he expected from the profession, and it had compromised his enjoyment of the work:

> But since a few years, with all this… what I mentioned a few times, all these extra *things*… I think 'shouldn't we get on with teaching, and leave all these other *things* outside the classroom?'
>
> (Matthijs, group 5, school 4)

A Shifting Sense of Professionalism?

Beyond teachers' experiences of performativity and the responses this incites, it is pertinent to consider the impact that is had on teachers' professional identities – their sense of professionalism – and the ways

respondents may have (re)oriented their beliefs and expectations to (better) align with the daily realities of being a teaching professional today.

One striking way the data indicates this might be happening, is through teachers' experience of autonomy, traditionally considered a key characteristic of the professions. Accounts indicate that traditional notions of professional autonomy may be shifting: many respondents understand it as something constricted, to be earned rather than a given. This was commonly described as autonomy 'between the lines.'

> *Robbie*: I feel like I am in control of what I would like to do. It's only that the IB'er [member of management] and colleagues, they help me to be on the red line, the main path, but which way I would like to go – it's up to me. As long as I achieve my goals that I talked about with management.
>
> *Interviewer*: And goals in terms of…
>
> *Robbie*: Numbers. Purely results.
>
> (Robbie, group 5, school 2)

When asked what would happen if these goals were not met, Robbie's response illustrates the creeping normalisation of the paramountcy of student achievement.

> … so, I'd have more observations in the class. So as soon as that happens then, I hope, to be honest, *I hope,* there would be more focus on me than there is right now.

Only a minority of respondents reported to have a meaningful role in choosing and shaping curricula – connected to their school's pedagogical and managerial approach (Browes forth.) – and yet, most, reported to be content with the autonomy afforded to them. For some, namely those 'aligned' teachers such as Robbie, these conditioned 'micro-autonomous spaces' (Wilkins 2011) were sufficient for them to consider themselves autonomous professionals. For others, this stance was merely practical; there was not the time for utilising the luxury of more meaningful curricular freedom. Indeed, teachers' inabilities to teach lessons with the desired levels of creativity and inspiration, as reported earlier, is generally not attributed to a lack of autonomy, rather, to a lack of time. In environments where time is scarce and tasks are many, we should not assume that curricular autonomy is desired. While often described as 'possible theoretically,' teachers do not have the time to spend developing curricula (unless this is an integral part of school culture), and many claimed to prefer to work from the textbook. Those who *did* desire more curricular autonomy were almost exclusively teachers who have been in the profession a decade or more;

> I really feel that it's too little. […]. We have a method for every single subject.
>
> (Emma, group 6, school 1)

Tight working schedules have other impacts too. Collegiality (in terms of taking time to discuss, share ideas, and support colleagues) may now also be take low precedence;

> Sometimes after work I am like: I *could* go to a colleague, but if I want to ask a question I know I will be there for ten minutes, because we start talking about other things while I also want to finish it [other work], and I want to leave at 4 o' clock for example.
>
> (Nick, group 6, school 3)

Finally, it is important to re-stress that the aim of recent reforms has been the *professionalization* of teachers. In the Netherlands, this has included the development of teacher competency standards and professional development courses aligned with these standards. These measures incite changes to the profession by (re)defining its parameters and what it means to be a good teacher. Respondents' experiences of this are mixed: several (predominately, but not exclusively, younger) teachers see the teaching standards and courses on offer as an opportunity for professional growth, others generally more advanced in their careers, see the initiative as tokenistic, externally controlled, and narrowly-focused on standards-based learning.[25]

It is therefore perhaps accurate to describe teachers' experiences of their professionalism as divergent: separated into those who have (re)oriented their values, expectations, and practices to (better) fit their reality, and those who feel frustrated, holding onto previous experiences and more 'traditional' notions of what it means to be a teaching professional. Although a small data set, findings indicate that members of these groups can perhaps best (if not always) be identified by their years of teaching experience. Despite these differing perspectives and the daily compromises some teachers feel they are making, all respondents appear to be able to reconcile their beliefs and practices to some extent; claiming they are still, in some form, able to be the sort of teacher they want to be. Particularly for those who feel they are making significant compromises daily, this is due to expectation realignment, the ability to take refuge in small successes, and the hope that, in whatever way chosen, they are providing crucial support to their students.

> *Interviewer:* and do you feel that you *can* be this teacher?
>
> *Emma:* Yeah, at this point I can. But at other levels I can't be, I can't. The challenge is that you always somehow feel it's never going to be perfect – there's always rules and things that you just forgot and missed out on and it's the process of letting go and accepting it, and thinking 'well, that was a good day anyway.'
>
> (Emma, group 6, school 1)

[25] These competency standards have since been rescinded.

> … you hope to make a difference, you hope that someday the kids will look back and maybe think of *me,* not about the math I explained, but as a person, as a teacher who listened and was there for them.
>
> (Lynn, group 8, school 5)

DISCUSSION & CONCLUDING THOUGHTS

This study has taken a look at a group of professionals operating within environments of surveillance, whereby they are increasingly held accountable through the recording of their work and their performance outcomes. This has been through a focus on primary school teachers in the Netherlands, using the concept of role discrepancy to understand what teachers do and how they feel about their work. In doing so, while situating professional change as an impact of broader, educational reform, it also demonstrates at the micro level, how the teaching profession might be changing from the inside, through the manner in which teachers' perceive their roles and responsibilities (Ball 2016).

Findings suggest that primary teachers in the autonomous Dutch system share many experiences with those working in notoriously high-stakes, highly-controlled PBA contexts. Respondents were found to be operating in a 'post-professional' environment (Hargreaves 2000), struggling under time pressures and overburdened with tasks, directly impacting what and how they taught (see Berryhill, Linney & Fromewick, 2009). The impact of this on teaching however, was reported to be less for those teachers working in schools that adopt project-based learning. Nonetheless, in all schools, tasks traditionally at the core of teaching are apparently being pushed out by tasks that, instead, have the function of *measuring* teaching and learning. In these work environments, as in other professions, role discrepancy was found to be prevalent. As found amongst social workers (Agresta 2006) and occupational therapists (Lloyd et al 2004), teachers desired to spend more time on their areas of specialism and less time on more generic, administrative tasks. Yet, role discrepancy does not necessarily result in feelings of frustration or compromise, rather, teachers were found to adjust to these realities and negotiate policy boundaries in quite different ways.

Teachers' responses to role discrepancy constitute either an adaptation of their work to allow space for beliefs & required performative practices, or regular compromises being made between the two. Whereas the first approach might enable teachers to minimise their experience of role discrepancy, this was not always in a desirable way. Teachers' perspectives and approaches to their work discussed in the paper should not be considered an exhaustive list, rather, were those identified amongst this study's sample. Neither should they be considered entirely exclusive. The majority of teaches described particular ways of working, yet a small number seem to adapt their approach as deemed necessary;

feeling the weight of task accumulation for example and thereby regularly working from home, while also at times 'evading' policy by avoiding particular performative tasks.

Significantly, a small minority of teachers, all early in their careers, did not appear to experience role discrepancy. Subscribing to the PBA agenda, they measured themselves as professionals against their students' performance outcomes and, despite some frustrations, did not feel the need to compromise or significantly adjust their practices to accommodate their professional values. In this sense, these teachers display the characteristics of Wilkin's (2010) 'post-performative' teachers, 'knowing' themselves as professionals through performative measures (Holloway & Brass 2018).

These teachers' formative professional years took place before key PBA reforms had (fully) taken hold, and their values and expectations were shaped within different environments (see Holloway & Brass 2018). Their experiences contrast with their more experienced colleagues (in the profession for at least a decade), who were more likely to feel the burden of compromise and a greater sense of frustration with their work. This finding is contrary to the hypothesis presented in Takase et al's (2006a) study on the nursing profession; that more experienced workers may experience role discrepancy to a lesser degree as they are *better able* to adjust or 'improve' their practice in line with their beliefs. Rather, teachers are not able to 'adjust' the multitude of demands upon them, and thus experience a sense of being pulled in two directions at once. Indeed, it is for this reason that the term policy *evasion* is used to describe those respondents who do, essentially, try to 'improve' their practice. As discussed by Perryman et al (2011), performative policies have become woven into the fabric of teaching, and normalised to the extent that 'policy resistance' is not possible. Instead, teachers in this study were found to 'drop' particular performative tasks for largely practical reasons, and only if they had the support of management to do so.

These findings complement those from Dutch secondary education, with Hendrikx (2019), noting that teachers either compromise on their beliefs, their required practices, or their time. Yet importantly, they also reveal a complexity to policy enactment that not only signifies the importance of *compromise*, but also the importance of *adaptation*. It is clear that respondents' different perspectives towards PBA and their various responses to the challenge of role discrepancy, reinforces the conceptualisation of teachers as 'pragmatic' and 'eclectic' policy actors (Moore et al 2002): positioning their professionalism in a way that aligns with policy demands (principled pragmatism), feeling forced into 'resignation' and compromising on professional beliefs (contingent pragmatism), or finding themselves somewhere in the middle of this 'continuum,' and altering their practices to accommodate beliefs, by accumulating tasks, integrating them, or evading them.

From these various positions, teachers held particular views on key aspects of their professionalism. Many respondents were content with what appears to be constricted autonomy or 'autonomy between the lines.' Whether for pragmatic purposes or ideological ones, such findings point towards a re-

professionalisation, rather than a de-professionalisation of teachers (see Ball 2016; Noordegraaf 2013). Despite the small sample, when considered alongside studies such as Holloway & Brass (2018), Moore et al (2002) & Wilkins (2011), this study suggests that performative policies might be creating a divisive chasm. For those 'post-performative' teachers standing on the mainland who have either been professionally socialised within these realities or who have effectively adapted to them, PBA perhaps offers clear and useful professional guidance. For others, whose professional values were shaped in a different policy context, they may feel stranded, unsure of what their profession has become. Moore and Clarke's (2016) 'attachment to professionalism' concept is useful in understanding this potential disjuncture, and may explain empirical data showing why some of these 'stranded' teachers feel the need to resign to the managerial demands upon them. Further research in the Netherlands would be needed to test this idea.

Policy-makers should be weary here, and strive to narrow potential chasms rather than risk the (further) alienation of a large part of the teaching profession. Stubbornly-narrow and, now cemented, quality measurements appear to restrict teachers in the classroom, and unclear quality assurance frameworks have led to the documentation of all aspects of their work and a task overload. In these ways, teachers' autonomy is constricted both directly and indirectly. Surveillance has its impacts: traditionally-autonomous professionals in a traditionally-autonomous system are being micro-managed through various layers of governance, and, based on shortage figures, it seems teaching has become an undesirable profession. Rather than continue struggling to save the profession through forced, external efforts to raise its status, attention should turn to the nature by which teachers are held accountable, and steps taken to support the development of accountability tools that are rooted within the profession itself, and based upon trust.

REFERENCES

Agresta, J. (2006). Job Satisfaction Among School Social Workers. *Journal of Social Service Research*, *33*(1), 15–26.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, *18*(2), 215–228.

Ball, S. J. (2016). Neoliberal education? Confronting the slouching beast. *Policy Futures in Education*, *14*(8).

Ball, S., Maguire, M., Braun, A., & Hoskins, K. (2011). Policy actors: Doing policy work in schools. Discourse: Studies in the Cultural Politics of Education. 32. 625-639

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.

Browes, N. (forth). Test-based Accountability & Perceived Pressure in an Autonomous Education System: Does School Performance Affect Teacher Experience? *Educational Assessment, Evaluation and Accountability*

Browes, N., & K Altinyelken, H. (2019). The instrumentation of test-based accountability in the autonomous Dutch system. *Journal of Education Policy*, 1-22.

Clement, M. and Vandenberghe, R. (2000). Teachers' professional development: a solitary or collegial (ad)venture? Teaching and teacher education, 16 (1), 81–101

Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research*, *37*(8), 677–692.

Day, C., & Smethem, L. (2009). The effects of reform: Have teachers really lost their sense of professionalism? *Journal of Educational Change*, *10*(2–3), 141–157.

Evetts, J. (2003). The Sociological Analysis of Professionalism. *International Sociology*, *18*(2), 395–415.

Evetts, J. (2009). New Professionalism and New Public Management: Changes, Continuities and Consequences. *Comparative Sociology*, *8*(2), 247–266.

Faulconbridge J. R. Muzio, D. (2012) The rescaling of the professions: towards a transnational sociology of the professions. International Sociology 27 (1) 109-125.

Hargreaves, A. (2000). Teachers and Teaching : Theory and Practice and Professional Learning Four Ages of Professionalism and Professional Learning. *Teachers and Teaching: History and Practice*, *6*(2), 37–41.

Hargreaves, A. (2005). Educational change takes ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education*, *21*(8), 967–983.

Hendrikx, W. (2019). What we should do Vs what we do: Teachers' professional identity in a context of managerial reform. *Educational Studies*, 1-17.

Holloway, J. & Brass, J. (2018) Making accountable teachers: the terrors and pleasures of performativity, Journal of Education Policy, 33:3, 361-382

Hoyle, E. (2001). Teaching: prestige, status and esteem. Educational Management Administration and Leadership, 29 (2) 139-159

Inspectie van het Onderwijs. (2020). Onderzoekskader 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs.

Inspectorate of Education. (2019). *State of Education, 2020, Summary Report.* Utrecht: Inspectie van het onderwijs

Lloyd, C., King, R., & McKenna, K. (2004). Actual and Preferred Work Activities of Mental Health Occupational Therapists: Congruence or Discrepancy? *British Journal of Occupational Therapy*, *67*(4), 167–175.

Moore, A., Edwards, G., Halpin, D., & George, R. (2002). Compliance, Resistance and Pragmatism: The (re)construction of schoolteacher identities in a period of intensive educational reform. *British Educational Research Journal*, *28*(4), 551–565.

Moore, A., & Clarke, M. (2016). 'Cruel optimism': teacher attachment to professionalism in an era of performativity. *Journal of Education Policy*, *31*(5), 666–677.

Muzio, D., Brock, D. M., & Suddaby, R. (2013). Professions and institutional change: Towards an institutionalist sociology of the professions. *Journal of Management Studies*, *50*(5), 699–721.

Demirkasimoğlu, N. (2010). Defining "teacher professionalism" from different perspectives. *Procedia - Social and Behavioral Sciences*, *9*, 2047–2051.

Noordegraaf, M. (2013). Reconfiguring Professional Work. *Administration & Society*, *48*(7), 783–810.

Nusche, D., Braun, H., Halász, G., & Santiago, P. (2014), OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing.

OECD. (2018). Country Note: Netherland. *Results from TALIS 2018.*

OECD. (2019). Country Note: Netherlands. *Education at a Glance: OECD Indicators*.

Olivant, K. F. (2015). I am not a format: Teachers experiences with fostering creativity in the era of accountability. *Journal of Research in Childhood Education*, *29*(1), 115–129.

Perryman, J., Ball, S., 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

Sachs, J. (2001). Teacher professional identity: Competing discourses, competing outcomes. Journal of Education Policy, 16(2), 149–161

Skerritt, C. (2020). School autonomy and the surveillance of teachers. *International Journal of Leadership in Education*, 1-28.

Strum, J., Groenendijk, L., Kruithof., B. & Rens, J. (1998) Educational Pluralism - a historical study of so-called 'pillarization' in the Netherlands, including a comparison with some developments in South African education, Comparative Education, 34:3, 281-297.

Takase, M., Maude, P., & Manias, E. (2006a). Role discrepancy: Is it a common problem among nurses? *Journal of Advanced Nursing*, *54*(6), 751–759.

Takase, M., Maude, P., & Manias, E. (2006b). The impact of role discrepancy on nurses' intention to quit their jobs. *Journal of Clinical Nursing*, *15*(9), 1071–1080.

Valli, L., & Buese, D. (2007). *The changing roles of teachers in an era of high-stakes accountability. American Educational Research Journal* (Vol. 44).

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

Verger, A., & Parcerisa, L. (2017). A Difficult Relationship Accountability Policies and Teachers : International Evidence and Key Premises for Future Research. *International Handbook of Teacher Quality and Policy*, *Verger*, *A.*(January), 241–254.

Waslander, S., Hooge, E., & Drewes, T. (2016). Steering dynamics in the Dutch education system. *European Journal of Education*, *51*(4), 478-494.

Wilkins, C. (2011). Professionalism and the post-performative teacher: New teachers reflect on autonomy and accountability in the English school system. *Professional Development in Education*, *37*(3), 389–409.

Yin, R.K., (1984). Case Study Research: Design and Methods. Beverly Hills, Calif: Sage Publications.

CHAPTER 5

POLICY ENACTMENT II:
What Role Does School Performance Play in Teachers' Experience of Pressure?

This chapter is a pre-print version of: Natalie Browes (forth). Test-based Accountability & Perceived Pressure in an Autonomous Education System: Does School Performance Affect Teacher Experience? Published online on June 21st 2021, in Educational Assessment, Evaluation and Accountability.
DOI: 10.1007/s11092-021-09365-9

INTRODUCTION

With schools worldwide being held accountable through their student test results, education has become a matter of performativity. Teachers are expected to align their teaching to core learning standards, demonstrate data-oriented working, and show improvements in their students' test scores. Holding actors accountable for education quality in this way is known as 'test-based accountability' or TBA. Research, particularly in Anglo-Saxon contexts, has shown such results-oriented policies to have exerted considerable pressures on teachers: not only changing what is required of them, but changing what it means to be a good teacher and the very identity of the profession (Ball 2003; Day 2002; Valli & Buese 2007).

A large number of studies have detailed the impact of (high-stakes) TBA policies on teachers; (Berryhill, Linney, & Fromewick 2009; Nathaniel, Pendergast, Segool, Saeki & Ryan 2016; Saha & Dworkin 2009; Verger & Parcerisa 2017; Wilkins 2011, to name a few) resulting in notorious maxims such as '…the terrors of performativity' (Ball 2003) and '(life in) the pressure cooker' (Agrey 2004; Perryman, Ball, Maguire, & Braun 2011). A far smaller number have directly compared the experiences of teachers in schools that occupy different positions within the same accountability system (DeBray 2000; Diamond & Spillane 2004; Mintrop 2007). This study adds to this scant literature, examining teacher experience of TBA and performative pressure in schools under the same management, yet that face different circumstances. This is done in the context of the Netherlands: a system where school (board) autonomy is high, parents have substantial choice in a diverse market of government-funded schools, and the

Inspectorate has a longstanding quality assurance role. The paper analyses two ‘high’ and two ‘low’ performing primary schools under the same board in one city. Dutch primary schools are particularly useful sites to research the impacts of TBA given the significant role of compulsory standardised testing throughout. These tests are used by the government to measure student and school performance and often, though more informally, used by boards and schools to gauge teacher performance. Drawing on the analysis of interviews with teachers, principals, and the school board as well as key school documents, this research seeks to uncover cultural and contextual elements of the schools, to understand their interpretation and enactment of TBA, conceived through the concept ‘logics of action,’ and, by extension, the performance pressures facing their teachers. The study highlights similarities and differences between teachers’ experiences, seeking to answer why these might exist and to what extent they can be understood through school performance level. It is shaped by two main research questions:

I. What accountability pressures do high and low performing schools face, and what logics of action are adopted?
II. Do teachers in high and low performing schools experience pressure differently: in what ways do they feel under pressure to perform, and what do they perceive to be the main sources of this pressure?

This research adds to our knowledge of the enactment and impact of TBA policies in various ways. Firstly, the field is dominated by studies conducted in Anglo-Saxon, ‘high-stakes’ contexts; that is, contexts where sanctions or rewards are attached to test performance, often at the school *and* teacher level. By researching enactment in an accountability system that does not implement such measures, the paper adds to the limited research (see for example Thiel & Bellmann 2017) that questions implicit assumptions that lower stakes mean no or limited undesired impacts. Secondly, previous studies have tended to focus on particular accountability measures and responses at the *school level*; the state imposed administrative sanctions facing probation schools for example (Mintrop 2004), or the market pressures facing schools in competitive systems (Van Zanten 2009). By *also* focusing on the individual experiences of teachers and how they perceive pressure, the study investigates whether fewer (external) performative demands on schools does in fact result in reduced pressure felt by teachers. Finally, by directly comparing teachers’ experiences of TBA in different school settings, the research attempts to highlight assumptions on which these policies are based, providing a deeper understanding of policy enactment and bringing into question ‘what works, where and for whom?’ The research forms part of a wider, comparative project called ‘ReformEd’ (reformedproject.eu), which investigates the evolution and enactment of TBA in various countries around the world.

TBA, Performativity & Enactment Theory

Comparable test results are central to the TBA model. Tests are designed to measure the attainment of centralised standards and used to gauge student and school performance. Attaching these test results to sanctions or rewards (managed at the municipal, state, or national level) is often referred to *administrative accountability* (see Verger, Parcerisa & Fontdevila 2019). The school inspectorate or equivalent authority, may also play a central role here; administering quality labels and interventions or even closures should a school not meet performance targets. These interventions, or the threat of them, place particular pressure on those schools on the borderline of these targets.

Working hand in hand with these administrative measures described above, the market has also been leveraged as an accountability mechanism. *Market accountability,* usually involving the publication of a school's test results and the Inspectorate's quality labels which are meant to guide parents when it comes to school choice, promotes competition between schools (Verger et al 2019). Underperforming schools may struggle to attract and/or keep students and may even have to close. A sufficiently-performing school that does not adequately diversify and innovate may also struggle to attract students (OECD 2010; OECD 2013). In these competitive environments, schools' reputations and the pressure of maintaining them, become particularly important (West, Mattei & Roberts 2011).

The publication and availability of results also serves to inform the parents of current students about the schools' performance and their child's progress. This establishes parents as (another) forum by which schools are held to account (see Bovens 2007). West el al (2011) refer to this as *participative accountability*. This is particularly significant when tests are high stakes for students; holding a certification, streaming or selection function (see Verger et al 2019). In this regard, schools are also accountable to their students (West et al 2011).

These policy tools are designed to ensure that teachers are focused on results and the job of results improvement; whether their work is formally assessed through student test scores or not. Studies investigating the enactment of these policies often employ the term 'performativity' (Ball 2003) to denote generally the changes to teachers' practices and identities, imposed and encouraged, through performance-focused, test-based school accountability policies. The term, used throughout this paper, does therefore not only refer to the process of testing itself, but to the accoutrements of TBA. Many enactment-based studies have highlighted the burden that performativity has placed on teachers:

> The spiralling demands of government initiatives, incessant record-keeping, education plans, targeting and inspections, have left teachers reeling
>
> (Mathison & Freeman 2006, p43)

Given the centrality of standardised test results within accountability systems, school performance level is a particularly important factor to consider when examining the pressures of performativity and the enactment of TBA. Enactment theory has underlined the complex dynamics involved in 'doing policy;' understanding school actors as *policy shapers*, pivotal in their interpretation and translation roles (Ball, Maguire, Braun & Hoskins 2011). It has also underscored time and again the utmost importance of context (see for example, Braun, Ball, Maguire, & Hoskins 2011): to understand enactment we must understand something about performance context (school competition, reputation, student population…), as well as school (performance) culture (Ball & Maroy 2009). These are fundamental factors which both shape and reflect policy translation. The extent to which a school's image is tied into academic performance is key here. Scholars' work on schools' 'logics of action', provides an excellent perspective from which to better understand this notion of performance culture.

Ball & Maroy (2009) define logics of action as consistencies in practices and decisions regarding various aspects of a school's external and internal functioning, including; student enrolment, curricular content, organisation of classes and performance. Linking to earlier work (Bernstein 1996; Hargreaves 1995), these practices or 'functions' are thought to exist within two domains: schools must "deal with an instrumental function, or task achievement, but also with an expressive function, or maintaining good social relationships" (Hargreaves 1995, p25). According to Hargreaves, these two domains are at the core of school culture.

Schools with a predominately *instrumental order* may appear more exclusive, with a high level of student homogeneity and an emphasis on academic performance and results maximisation. Ability differentiation and programmes that target high-performing students are common and hierarchy and authority are likely to play a strong role in shaping the relationships between students and teachers as well as teachers and management. Conversely, schools where the *expressive order* is dominant can be recognised through the importance attributed to inclusivity and equity, often concentrating on low performing students and emphasising student socio-emotional development. Interpersonal relationships play a central role in the organisation of the school, including more open relationships between teachers and management and greater collaboration and teacher involvement in decision making (Ball and Maroy 2009).

These functions can be understood as a response to regulatory processes (accountability structures for example), as well as to more local processes (the nature of and the school's position within the education market) (Maroy and van Zanten 2009). They serve to promote a particular 'image,' to attract and retain a certain profile of student. We would therefore expect to find a relationship between a school's logics of action, student population, performance, and market position. An understanding of these factors is fundamental to understand the impact of accountability measures on schools, and by extension, the expectations on teachers and the pressures they experience.

State of the Art

A number of studies have looked at the enactment and impact of TBA, mediated by school performance within a system (Diamond & Spillane 2004; Mathison & Freeman 2006; Mintrop 2004), the nature of the accountability system; 'high' or 'low' stakes (Thiel & Bellmann 2017; Nathaniel et al 2016), and evolution along the TBA timeline (Holloway & Brass 2018). Comparing high and low performing schools in two 'high-stakes' American states, Diamond & Spillane (2004) reported particular pressures and undesirable responses in underperforming probation schools. These schools were generally attended by high percentages of poor and ethnic minority students. The studies found that school responses to accountability sanctions focused on superficially complying with policy demands (Diamond and Spillane 2004) resulting in narrowed curricula and educational triage, whereas high-performing schools implemented more comprehensive and meaningful changes that were 'closer to the intentions of policy-makers' p.1157). Significantly, Thiel & Bellmann (2017) revealed that undesirable side effects are not only present in schools operating in high-stakes contexts, but also in 'low' and 'no- stakes' contexts. Indeed, even in contexts with no formal administrative sanctions or free school choice, the fact that test scores have become inherently tied to perceptions of school and teacher quality, exacerbated by the publication of these scores, may result in considerable reputational pressures.

At the teacher level, empirical studies such as those conducted in the English and US contexts have found (high-stakes) accountability to negatively affect the curricular and pedagogical freedom of teachers; whether they are explicitly required to follow increasingly-detailed prescribed curricula (Dobbins 2009; Webb et al 2004) or manipulated into narrowing curricula due to the stakes attached to tests (Berryhill et al 2009; Mathison & Freeman 2006). In a particularly comprehensive paper, Valli & Buese (2007) reported that teachers in one US mid-Atlantic state experienced a loss of control over the curriculum in terms of its pace, content and organisation. Mathison and Freeman (conducting research in the state of New York) found that, as a result of such performative pressures, teacher stress was prevalent across schools, regardless of their performance level.

Within these neoliberal systems of accountability and surveillance, core aspects of the teaching profession appear to be changing. *Regulated autonomy* for example (Dale 1982), describes the existence of teachers' autonomy within a limited scope, constrained by increasing external control. Perryman et al (2011) present the notion of *earned autonomy*: awarded and dependant on (continued) good results. Such change is not only evident in teachers' work environments, but also in their expectations and perceptions of the profession. Guided by Ball's work (2003; 2016), Holloway & Brass' findings (2018) reflect the tightening grip that TBA is having on the teaching profession. Comparing two separate studies, US teachers interviewed during the 'second wave' of TBA (Obama's 'Race to the Top') expressed markedly different views from those interviewed almost a decade earlier amidst Bush's 'No Child Left Behind' Act. Whereas teachers experiencing TBA in its early stages saw it as an external

intrusion with clearly negative impacts on their autonomy, practices, and professionalism, ten years later these policy tools had apparently become woven into the fabric of teaching. Not only were test data, standards, and performance indicators valued, but TBA mechanisms had become 'the very modes by which they (teachers) knew themselves and their quality' (Holloway & Brass 2018 p 362). In such environments, with quality reduced to measurable output and value placed on recording work over the work itself (see Ball 2016, Day 2002), it has been argued that a professional teacher is more akin to a technician; expected to 'do things right' rather than 'do the right thing' (Darling-Hammond 1990 p31). How these issues play out in a system such as the Dutch one, characterised by its high levels of decentralisation and school and teacher 'freedoms,' needs to be explored further.

Accountability in Dutch Schools

In Dutch primary schools, centralised core learning standards are measured primarily through the final or 'end-test' (eindtoets) which takes place in group 8, the final year of primary school (akin to grade 6). Schools must also be able to show they are following the development of students in relation to these standards through a student tracking system 'leerling volg systeem' or 'LVS.' This system is test-based, with tests usually taking place twice each year and compulsory as of group 3 (when students are 6-7 years old), although many schools already start them in groups 1 or 2. There are a number of companies that provide the end-test and LVS tests, and schools can choose which to opt for.

Beyond these learning objectives, there is no core curriculum that schools must follow, however, spotting an opportunity, a number of private companies have developed subject curricula, known as 'methods' in line with learning standards and test content. These 'methods' are highly-structured and detailed, and often employ end-of-unit testing. As part of the tracking culture, various student administrative systems, also developed privately, are utilised by schools. With these, teachers are encouraged to record and analyse student results, develop 'class' or 'group plans' (the results-based ability grouping of students and subsequent curricular plans), and keep (daily) notes on students' work, progress, and behaviour, as well as a record of all parental interactions. Using these teaching methods and administrative systems is not compulsory, although the Inspectorate states that schools must be able to show they are following the development of their students in a systematic way (Inspectie van het Onderwijs 2020).

Through the standardised tests, Dutch primary schools are held accountable in a number of ways. If a school performs below average in the end-test for three consecutive years[26] it is labelled as 'very weak' by the Inspectorate and undergoes an intensive inspection and monitoring process whereby a selection of other standardised test results and education processes will also be examined. Schools must produce

[26] Compared to schools with a similar SES.

improvement plans and are encouraged to seek external support from one of a number of providers to help get performance 'back on track' (Inspectie van het Onderwijs 2020). If a school consistently underperforms, showing little or no sign of improvement, the Minister of Education has the right to withdraw funding, effectively closing the school (Waslander 2010). However, policy experts – interviewed at an earlier stage of the research – could only recall this happening on 'one or two' occasions (Browes & Altinyelken 2019). Teachers' salaries and bonuses are not connected to test scores and neither are teachers *formally* assessment through scores[27] however, administrative requirements are powerful in shaping the priorities of schoolboards and principals, and by extension, shaping teachers' practices.

Market accountability mechanisms work alongside the administrative. In the Netherlands, national and municipal-level websites provide easy-to-read, school-level data (see vensters.nl or allecijfers.nl/basisscholen). This includes some qualitative data (such as school plans), but is mainly quantitative (test scores and secondary school advice). Performance data is shown against national averages and can also be compared directly with other schools. Given that funding is attached to the student, based on parental education level, a poor-performing school may not only suffer from a reduction in student numbers but also a reduction in money. To encourage market dynamics, satisfactory-performing schools may now also request a quality inspection. Here, the Inspectorate assesses various aspects of a school's organisation and functioning, with the view of awarding a label of 'good' or even 'excellent' (see Inspectie van het Onderwijs 2020).

TBA in the Netherlands therefore differs in important ways from TBA in high-stakes systems, such as those found across the US. According to Thiel and Bellman (2017), given the lack of formal (and utilised) sanctions and rewards attached to test performance, the Netherlands could be classed as a 'low-stakes' system. Yet the looming presence of the Inspectorate and the publication of performance indicators provide a constant reminder to school actors of the need to play the performance game, and teacher stress and burnout are significant (Inspectorate of Education 2019). The very fact that results are published in a market where parents exercise school choice, makes a 'low-stakes' label problematic. Beyond this, given their role in secondary school streaming, standardised tests are high stakes for students, particularly in the upper years (groups 6, 7 and 8). Dutch secondary education varies considerably, ranging from more practically-focused schools 'voorbereidend middelbaar beroepsonderwijs' or 'VMBO' (leading to vocational training upon graduation), to 'hoger algemeen voortgezet onderwijs or 'HAVO' (leading to university of applied science), to 'voorbereidend wetenschappelijk onderwijs' or 'VWO' (theoretical education leading to research university) (see Nusche et al 2014). Movement between streams is challenging, making performance at primary school

[27] Indeed, there is no centralised framework for teacher evaluation, rather this is left up to the schools and boards to decide.

a core determinant of access to higher education and future career path. This creates an environment in which participative accountability may also play a significant role (see for example CNV Onderwijs & EenVandaag 2018; Duo Onderwijsonderzoek & De Monitor 2016). In summary, the current definitions and binary 'high-stakes', 'low-stakes' categorisations of accountability systems are limiting. For the reasons outlined above, it is perhaps most accurate to describe Dutch accountability as a 'middle-stakes' system.

Therefore, despite important differences regarding how accountability operates, Dutch teachers are regulated in many of the same ways as teachers in high-stakes contexts. An international comparison of teacher professionalism structures (Voisin & Dumay 2020) indicates that the Dutch system can be best understood as a 'market model:' market and standards-based regulation, characterised by, amongst other things, relatively low levels of professional autonomy (p.7). This model is also found in England and the US (amongst others). Given the high rates of school autonomy that define the Dutch system, this may seem counterintuitive. Indeed, considering the numerous intermediary stakeholders involved in the system and the numerous ways in which schools are driven to improve, it is important to examine these mechanisms and their effects on teachers more carefully.

METHODS

The research adopts an exploratory case study approach (Yin 2009). It compares the experiences of teachers across four primary schools, differing in their performance levels. As such, can best understood as a multiple case, embedded approach (Yin 2009). Purposive sampling (Patton 2015) was used to select schools, based on the criteria of school management, school 'type,' school location, and school performance level.

Schools under the same board were selected to keep external management conditions as equal as possible. All of the schools are public ('openbaar'), meaning they are government-established[28] as well as government-funded, and cannot discriminate or tailor education based on religious or philosophical grounds. In total, the board manages 26 schools, 21 of which are general primary schools. The schools are located in a small but densely populated city in the province of North Holland. It was necessary to select an urban area with a competitive education market, given the study's interest in the role of the market in accountability and performance pressure. Big cities were avoided for fear of research fatigue, indicated during an initial exploration of the field. The researcher created an excel sheet of all schools under the board, 'ranking' them based (only) on performance data, and contacting two of the top and two of the bottom ranked schools for participation in the study. Given the focus of the research,

[28] However, one of the schools was created through a merger of a public and a general independent school.

'performance' was determined in a very narrow, quantitative way, based on; 1) the schools' average scores in the end-test, and 2) the percentage of students going to VWO (the most academic level of secondary education). Beyond these figures, the term does not reflect school or teacher quality and says nothing about student, parent or inspectorate satisfaction. It is noteworthy that while both of the 'low performing' schools serve highly-mixed and non-ethnic Dutch populations, the 'high performing' schools are attended by almost exclusively middle class, ethnic Dutch populations. This high correspondence between socio-economic composition and school performance is not coincidental (Ball & Maroy 2009; Diamond & Spillane 2004).

In each school, interviews were sought with the principal and four teachers, including: one group 8 teacher (the 'end-test year'), another teacher from the 'bovenbouw' (groups 6-8, where tests are more high stakes for schools and students), and two group teachers from other tested years. Teacher recruitment for the research was challenging for various reasons, resulting in an incomplete sample and unavoidable instances of selection bias. In all schools the researcher first made contact with the principal to explain the study and request participation. Principals were concerned about adding to the high workloads of their teachers and would not enforce participation, rather agreed to inform teachers should any show interest. Akin to voluntary response sampling, or self-selection bias (Collier & Mahoney 1996), it is worth noting that these participants may be particularly politically engaged, or have particularly strong opinions about the policies under investigation. In case there were still gaps in the sample, 'missing' teachers were contacted by the researcher. To encourage participation and show appreciation, all interviewees were presented with a book voucher upon completion of their interview. In one school, neither of the two group 8 teachers agreed to participate, bringing the number of teacher interviews to three. However, this included a teacher who had recently taught group 8. In another school, the vice-principal stood in for the principal who was unavailable for an interview. This brought the interview total to 20; school board (n=1), (vice)principal (n=4), teacher (n=15). For a full list of participants, see table 4 at the end of this section. All schools and participants have been given pseudonyms to maintain their anonymity.

Interviews were conducted one-to-one, lasted between 32 and 57 minutes and were semi-structured. The interviewer followed a script but remained flexible to pursue relevant issues that were raised. Interview scripts differed slightly between actors, yet contained the same themes and many overlapping questions so that data could be triangulated. Themes included: teachers' work routines and working environments, testing practices and opinions on TBA, and experience of performance pressures and autonomy. Interviews were transcribed verbatim and then analysed using 'Atlas.ti'. To gain an insight into schools' logics of action, the most recently available school guide was downloaded and also coded. Schools are required by law to produce this document annually. It provides information to parents of prospective and current students about the mission, vision, working methods, education plans and

performance of the school. The comprehensive nature of these documents and their function to inform but also promote the school, make them an excellent data source. Content analysis, using a directed approach (Hsieh & Shannon 2005) was conducted on both the interview scripts and the school documents. For this purpose, a code list was developed iteratively, guided by theory and the research questions. Examples of codes include: school context_perceived reputation, policy interpretation_opinion on TBA, policy translation_test preparation, policy translation_ student differentiation, perceived pressure _inspectorate, perceived pressure_parents. The interview and coding scripts were developed with fellow ReformEd researchers as comparative research tools, but allowed for the addition of case-specific issues.

Table 4: List of Participants

Organisation	Name	Role
School board	Thomas	Director
Barlaeusschool	Yvonne	Vice-principal
	Eva	Teacher: group 8
	Daan	Teacher: group 7/8
	Sara	Teacher: group 6
	Marianne	Teacher: group 3
Basisschool De Witt	Bas	Principal
	Thijs	Teacher: group 8
	Ilse	Teacher: group 8/behaviour coach
	Anna	Teacher: group7
	Finn	Teacher: group 5
Erasmusschool	Monique	Principal
	Mark	Teacher: group 7
	Felix	Teacher: group 6
	Els	Teacher: group 2
Van Schurmanschool	Annemarie	Principal
	Saskia	Teacher: group 8
	Thea	Teacher: group 7/8 (SBO)
	Marieke	Teacher: group 6
	Jonathan	Teacher: group 5

FINDINGS

Findings are presented according to the study's research questions. The first section focuses on the school level: examining school contextual and cultural factors to try and determine the importance of performance and the degree of performance pressure in each school. The next section shifts to the teacher level, to see how this impacts teachers' experiences of performance pressure, and if and how this experience varies between schools.

Performance Context & Logics of Actions: A Study of Four Schools

The table below presents the contexts of each school, including their population, performance, and market position as perceived by the school's (vice) principal. This is followed by a more in-depth qualitative analysis of the schools, focusing on; accountability pressures, school values, and key school practices such as student recruitment and the organisation of learning.

Table 5: School Characteristics[29]

	Perfor-mance level	No. students (2019/20)	% 'extra funding' students (2018/19)*	% 'NOAT'** students (2019/20)	Av. end-test scores (2017,18,19)	% attending VWO (2017,18,19)	Inspection status	Enrolment & perceived competition
Barlaeus-school	'High'	432	0%	1%	(IEP) *** 88.6, 85, 86.9	47, 49, 50	'Without current judgement'	Recent decline in enrolment. A moderate-strong amount of competition perceived
Basisschool De Witt	'High'	603	0%	3%	(CITO)**** 542.1, 541, 539,6	68, 55, 49	'Good'	The school has a long waiting list. A sense of competition is not felt.
Erasmus-school	'Low'	170	35.5%	88%	(CITO) **** 525.2, 531, 528.6	12, 4, 14	'Good'	Sometimes struggles to achieve desirable student numbers. Not perceived to be in competition with other schools
Van Schurman-school	'Low'	234	22.5%	58%	(CITO)**** 522, 528, (IEP)*** 87.7	4, 7, 19	'Very weak'	The school has a waiting list. A moderate amount of competition is felt.

*2018/19 is the last year this data is available

** 'NOAT' stands for 'non-native speakers in Dutch education'

*** IEP scores range from 58 to 97. 25% of schools score <78 points, 25% score >84

**** CITO scores range from 513 to 548. 25% of schools score <532, 25% score >537

High Performing Schools

Barlaeusschool and Basisschool De Witt are amongst the top performing schools under the board. Located in affluent parts of the city, no students receive extra 'disadvantage' funding and percentages of NOAT students are very low. Parents at De Witt are described as highly educated, wealthy, working professionals, while Barlaeus' population is reportedly somewhat lower-middle class, and increasingly

[29] Schools are listed in no particular order

consists of blue-collar workers. As will be presented below, the schools' organisational values and their organisation and content of education reveal that an 'instrumental order' dominates in the schools.

Given that both schools comfortably exceed quality requirements, pressures concerning potential administrative sanctions are low. As the result of sufficient performance, it had been some time since Barlaeus received a school inspection (status: 'awaiting current judgement'). Conversely, at the time of research, De Witt had just completed an elected quality inspection. Following a short but intense inspection, the school was awarded the label of 'good.'

Despite the absence of any administrative interventions, pressure to maintain results was evident in principal interviews. This pressure was inherently connected to the schools upholding their academic reputations, and thus continuing to attract somewhat privileged and high-performing students. Both principals mentioned the high involvement and high expectations of parents. Bas, the principal of De Witt, further stressed that given his school's privileged population, the Inspectorate expects more than 'just sufficient' results. Despite this, he reported to feel confident in the positive and distinctive reputation of De Witt, and spoke proudly of its long waiting list. Beyond its academic success, Bas considered the school's warm and collaborative culture absolutely central to this reputation.

> The most different thing was that my policy is 'you work together'. Open the doors, speak with each other! Also with the parents we open the doors. Before I started, the parents bring their kids and must wait outside in the morning, and that was the first change – 'come inside!'
>
> (Bas, principal, Basisschool De Witt)

In contrast, maintaining sufficient student numbers was a considerable concern for Barlaeus' vice-principal, Yvonne, an issue that had only arisen in recent years. This appears to be the result of a change in market position: on the one hand a newly-opened school in close proximity had increased competition for students, while on the other, a reportedly changing neighbourhood composition had led to a perceived mismatch between the school's logics of action and the demands of local families.

> "…now we want to do more of the nice things. More creativity, more music, every day from 12.00 until 14.00, so all the groups are working on nice things to do. It had to be more attractive, not simply lessons out of the book."
>
> (Yvonne, vice-principal, Barlaeusschool)

Certainly, for both of these high-performing schools, offering a broad, well-rounded educational experience where regular time is dedicated to gym, music and art, was of high importance. However, the core learning areas and academic achievement, it was stressed, remained a priority, and a performance orientation was evident at the schools. At Barlaeus, this manifested itself through differentiated working, a particularly strong focus on high-ability students, and an emphasis on results-oriented working (for example, participating in data-literacy projects with the explicit aim of increasing student outcomes). Teachers at the school are constructed first and foremost as individual professionals, encouraged to take the initiative to develop themselves and to work towards their own strengths in order to achieve desirable results. De Witt also promotes a strong emphasis on student achievement and a

differentiated or 'customized' approach to student education, yet one that also appears tailored to struggling students as well as to high achievers.

Teaching 'methods' were used at both schools, however, whereas curricula at Barlaeus appears to be (exclusively) prescribed, De Witt is now moving away from the use of purchased methods for non-core subjects. On this matter, De Witt's principal emphasised the importance of curricular autonomy to foster (teacher) innovation and 'avoid dependency,' yet, given that this has not been extended to the assessed learning areas, at least for now, performance seemingly takes priority over innovation. Both schools also implement LVS tests from group 2 onwards. These results are recorded in biannual student reports and discussed with parents. The data is discussed individually with management as well as in team meetings, and teachers' class plans, which are based largely on results, are regularly reviewed. In both schools, the importance of results 'justification' was also emphasised: a teacher experiencing a results 'dip' was admissible, as long as there was an acceptable explanation (i.e. one not linked to the role of the teacher). 'Unexplained' drops in results would be met with concern

> We have school results, which we can see from each other, and we are transparent about it. You can see how your group is doing compared to the other groups of the same age. And, it's not a problem when there are different scores, but it can be a problem when you can't explain it.
>
> (Bas, principal, Basisschool De Witt)

Similarly, according to Barlaeus' vice-principal, concern about 'unexplainable' results would be followed up with teacher meetings and a member of management observing classes to establish the cause of the performance decline.

Low Performing Schools

Erasmusschool and Van Schurmanschool are located in majority 'immigrant' and mixed neighbourhoods respectively, dynamics that are reflected in the schools' populations. Income levels in these neighbourhoods are below the city average and significant numbers of students at both schools receive extra funding based on the criterion that parental educational levels are low. At Erasmus in particular, many new students reportedly speak little or no Dutch, and in response, the school has forged itself a reputation as somewhat of a language specialist. Van Schurmanschool is attached to a school for special education (sharing a principal), with an aim of eventually integrating these students. As will be seen further, the dominance of an expressive order is evident in the schools' oft-discussed values, and organisational and educational dynamics.

The average performance of both schools has been amongst the lowest under the board. Their principals reported to feel the pressure of the Inspectorate's eye upon them, and considered their 'quality status' precarious given their student populations. Yet administrative interventions differ considerably between the two schools. Dipping below the average mark for its end-test results for three consecutive years, Van Schurman was labelled 'very weak' by the Inspectorate, and was undergoing a period of extensive inspection at the time of research. Classed as an underperforming school, measures included; bringing

in external support, monitoring from the Inspectorate, and the utilisation of a school improvement plan. Performing at or above average, Erasmusschool had received no such risk-based visit. The principal recounted its last (basic) inspection as a positive process and highlighted the favourable report it had received.

Despite relatively low performance, neither principal perceived their school to be under market pressure. While Erasmus' student numbers were reported to fluctuate slightly from year to year – sometimes making target enrolment numbers difficult to reach – this was not attributed to school competition but to a combination of family relocation (the result of housing developments elsewhere in the city), and a number of its students moving into special education. Conversely, Van Schurmanschool was reported to have a waiting list, with the principal attributing its popularity to its reputation as a safe and pleasant place with good teachers. This reputation was reportedly built through word-of-mouth amongst the Dutch-Moroccan community in particular. However, Annemarie, the school's principal, also acknowledged that many 'white Dutch parents' living nearby choose to send their children elsewhere, something she believes may have been exasperated by the school's classification as 'very weak'.

Unlike the more instrumental orientations of the high performing schools, at Erasmus and Van Schurman, educational needs are emphasized over educational outcomes. Throughout the school guides and principal interviews, the institutions emphasise first and foremost inclusivity, equality, respect and tolerance.

> And what we do is we really look – what is what does the child need? Every child. We always ask ourselves, 'are our children benefiting from this?' when we want to do something new.
>
> (Monique, principal, Erasmusschool)

As in the high performing schools, differentiated learning is integral to the structure of education, yet at Erasmus and Van Schurmanschool it is more designed to incorporate academically-struggling students than to push high achievers. Differentiated learning and support for lower achieving students is particularly significant at Van Schurman, which is working towards dividing learning into a 'theoretical' and an 'entrepreneurial' stream from group 6.

Unlike the broad educational offers at De Witt and Barlaeus, curricula at the two low-performing schools appears to be considerably narrower. Non-core subjects, while desirable, were seen as luxuries for which the schools had insufficient time. This feeling was most prevalent at Van Schurman. Annemarie emphasised her expectations that staff follow set guidelines and structured learning in order to 'stay on track', with the school recently switching to new methods in several areas. Teachers, she believed, still had autonomy within these guidelines:

> "If you go to teach in a free school, or in a Montessori school, then you teach in a different way than you teach in this school. But if you go here to school, we all write down the programme on the board, and we all write down what the children learn. We all work with an instruction table. But you are free how you do it."
>
> (Annemarie, principal, Van Schurmanschool)

In comparison, while these methods were equally important at Erasmus, teachers appeared to have a much greater role in their choosing. Monique, Erasmus' principal, reported an expectation that her teachers take the initiative to explore new ideas, whilst making decisions as part of a team.

> I expect them to take their own responsibility... I give them space for that and you can lift them up too and I expect them to commit themselves to the way we work - so we speak to each other together, and together we make the rules, so I expect them to keep the rules in mind.
>
> (Monique, principal, Erasmusschool)

As in the high-performing schools, LVS tests at Erasmus and Van Schurman are taken twice yearly from group 2, and results are central in discussions with parents, management, and amongst teachers. In contrast to the high performers however, testing is primarily valued as a useful way to identify student learning problems. While teachers in both schools are encouraged to share ideas and tips, and discuss reasons for well-performing and under-performing groups, it was stressed that these meetings are based on support rather than judgement. Since the school was deemed 'underperforming,' Van Schurman's principal explained that data analysis and keeping records of teaching and learning, are, more than ever, stressed as vital parts of her teachers' work. Classroom observations, she noted, are also much more common place.

The principals in the low performing schools do, however, adopt different perspectives when it comes to test data. Monique, principal of Erasmus, expressed a real interest in the data generated from tests, understanding it as a useful way to gauge the level and progress of a group and to assess the school's performance against others. In contrast, at Van Schurman, comparative test data was generally perceived negatively. This data was the reason behind the schools' 'very weak' quality label, an assessment that was seen by the principal as essentially unfair, and as having resulted in undesirable effects on student learning; namely, an overly-narrow curriculum.

> …It has changed because the Inspectorate has such a high norm (…) But it is not always good for children – they have to play, they have to draw, they have to make music. But I can't say to the parents 'your child can't count, but he can really dance lovely.
>
> (Annemarie, principal, Van Schurmanschool)

Teachers' Experiences of Performativity & Pressure

There are clear commonalities between teachers' experiences of TBA across the four schools. Reflecting the numerous documented experiences of teachers in the 'high stakes' Anglo-Saxon contexts presented earlier, interviews revealed that, regardless of their school's performance level and logics of action, teachers felt the weight and frustrations of a performative, outcomes-based agenda. This included, to various extents, a sense of restriction in the classroom: being contained by, and carried along with, a rigid schedule and prescribed curricula shaped by test content (see also Valli & Buese 2007). The vast majority of participants reported to be conscious of having to 'stay within the lines' when it came to the core learning areas: sticking closely to the subject methods used in the school. All

teachers reported to have more autonomy outside of these core areas. Further, regardless of school performance level, almost all respondents reported to engage in test preparation activities. The nature of these activities varied, although mostly, teachers reported to take time to 'familiarise' their students with the structure of test questions. Whereas in the high performing schools, test preparation activities were reported to take place only in the few weeks before the tests, teachers in the low-performing schools were more likely to integrate them as a routine part of their teaching throughout the year. Similar findings were recorded by Diamond & Spillane (2004) in the US context.

Drawing further parallels with high-stakes contexts (see for example, Mathison & Freeman 2006), heavy workloads and a sense of insufficient time for tasks was also experienced by teachers across the four schools. Participants expressed a desire to have *more* time: the time to explore new topics and revisit challenging ones, and the time and flexibility to follow intrigue and to teach impromptu. Teachers also expressed a wish to dedicate more time to non-core subjects and for more creative, exploratory, and student-centred learning. The majority of participants wanted to spend *less time* recording and planning their work, particularly in relation to the time spent on class plans and updating student administrative systems. Several teachers across the schools reported to regularly work on their evenings and weekends in order to manage their tasks.

In sum, regardless of school context and performance level, all teachers interviewed recognized themselves to be working in performative environments and were exposed to some degree of pressure around (test) performance (Diamond & Spillane 2004; Mathison and Freeman 2006). Yet perceptions of pressure, its source, and the extent to which it affected participants, varied. Some teachers found a performative work environment quite natural. These individuals considered testing to be a useful indicator of learning, and valued data and documentation for their ability to inform teaching.

> We put everything [the test data] in our system and we have an intern begeleider, and she looks with us… 'What did you do in your lessons?' 'What went wrong or what went well?' 'Why are the kids growing or 'what can we do to make them stronger next time?' So you look at the system and the system gives the answers.
>
> (Eva, group 8, Barlaeusschool)

In these cases, testing was seen as a clear way for teachers to know their goals, keep them focused, and gauge their own as well as their students' progress (see also Holloway & Brass 2018). As a result of these views, they were also more likely to report feeling pressure that their students perform well in tests. Another group of teachers were more sceptical, seeing tests as having a specific, yet limited function. These individuals spoke of their confidence in themselves as professionals, and did not connect their professionalism to their students' test scores. As such, most also claimed not to internalize performance pressure, although a small number of these teachers confessed they did, despite themselves, instinctively measure their teaching abilities through their students' results.

Data analysis did not reveal clear links between school performance level and the value that teachers attribute to testing. Rather, teachers spoke about the impact of their experience: those who had been in the job a decade or more were less likely to admit to attaching importance to results and were more likely to feel frustration with the growing centrality of testing and the role of data and learning standards within their work. These teachers reported to be witnessing a changing profession – a slow but steady tightening of the reigns.

> …in these developments in the last 10 years, we've become more and more slaves of the method – 'we need to do lesson 5 and then you need to go to this topic and then this…' there's very little room to do something for fun, to do something because it is just *interesting* but that doesn't add up to any measurable results.
>
> (Sara, group 6, Barlaeusschool)

Despite some adopting a more critical view, significantly, no teacher rejected testing outright. All respondents found value in test data. The majority of teachers claimed to appreciate the comparative function of standardized tests, not only accepting, but embracing the role of benchmarking. The ability to compare their class with a national average was seen as a particularly useful gauge of progress and for some, a reassurance of their teaching. Further, while it was acknowledged that a focus on the core standards detracted from other things, the learning goals measured by the tests were widely accepted. No teacher claimed to intrinsically value high test scores, yet all wanted their students to succeed in the system by achieving the results they knew them to be capable of.

Sources of Teacher Pressure

Unlike the individual variations seen in terms of the perceived importance of testing and the degree to which teachers internalized performance pressure, the impact of pressure and its perceived sources were found to be closely linked to school performance. These findings, explored below, further indicate that TBA pressure is not contained to high-stakes systems nor to low performing school contexts.

Administrative Accountability Teachers in both of the low performing schools perceived their school principal to be under performance pressure. This, they believed, came from an external source – the school inspectorate. Yet, the impact of this on teachers varied considerably between the two schools. Despite recognising the precariousness of their schools' situation in the face of administrative accountability measures, teachers at Erasmus reported their experience of performance pressure to be relatively low. The school was performing sufficiently in standardized tests and well overall, and teachers felt confident in their work and encouraged and supported by management. Teachers at Erasmus were also generally satisfied with their levels of professional freedom, and reported to have critical discussions as a team about the value of prescribed methods and the role of testing.

> Because some schools, they just do all the lessons, always follow all the lessons. And we already choose – 'no, just a few lessons are most important.'
>
> (Mark, group 7, Erasmusschool)

In contrast, managerial pressures at the 'under-performing' Van Schurmanschool had clearly filtered down to its teachers. Respondents from the school reported the greatest results-focus and performance pressure out of the four cases; 'it's [student results] the most important thing I think' (Saskia, group 8, Van Schurmanschool). All four teachers interviewed, and Saskia in particular given her responsibility as a group 8 teacher, found the ongoing inspection process to be a highly stressful affair. These teachers expressed greater feelings of insufficient autonomy than at the other schools and a sense of loss of control over teaching and tasks. Several felt that, in the face of administrative interventions, their school was pursuing a particularly narrow educational approach, and lacked a clear and collective vision, resulting in confused priorities and, at times, in ill-considered, ill-implemented working methods:

> …Instead of, thinking about it properly, they [the school management] implement all these things – so 'we try this, oh it doesn't work,' 'we try this, nope it doesn't work' (…) For me, that's not the right way to work, but okay, I try it, I don't try it fully, I just try it a little bit.
>
> (Jonathan, group 5, Van Schurmanschool)

This is comparable to the 'superficial' responses recorded by Diamond & Spillane (2004) in pressured probation schools in Chicago. Jonathan connected these pressures to his school's population. He believed his workload could be reduced and his overall professional satisfaction increased in a different school.

> [I'm] 75% happy. 25% thinking that there's more, that I'm working too hard for the money I get, the hours I make (…) That could increase at a different school. And that's probably because right now we have the inspection, and we have a lot of children who need extra attention, need the extra care, and that takes up a lot of time.
>
> (Jonathan, group 5, Van Schurmanschool)

More so than elsewhere, teachers at Van Schurman expressed a sense of unfairness about the accountability system: that the format of the standardized tests disadvantaged their non-native speaking students and the school was unfairly judged as 'very weak' before the inspector even arrived. Results were, as such, primarily attributed with instrumental value – perceived as a route back to greater autonomy and a more rounded school experience.

Conversely, the Inspectorate was not once mentioned as a source of results-based pressure at the high-performing schools. Nonetheless, teachers at De Witt did speak of the stress brought about by their recent inspection, despite its 'reward-based' focus (i.e. the reputational promotion of the school). This stress did not spring from test performance, but from a different aspect of the accountability agenda – documentation:

> The other week, we had the inspection, so we needed to know – 'is everything how it's supposed to be?' 'Is everything on paper?' 'Does it look right?'
>
> (Ilse, group 8, Basisschool De Witt)

Indeed, this quality inspection was reported as having a significant effect on teachers' working practices, with extra demands from management to ensure thorough accounting of their work should proof be required. Yet, whilst this form of administrative pressure did result in extra work, it did not appear to

impact teaching and teacher autonomy to the extent that it did in Van Schurmanschool. Rather, teachers at De Witt felt confident and clear in their school's educational approach and reported to have an important degree of freedom in their teaching, particularly in the non-core subjects.

> We have a lot of autonomy. 'Bas' is not watching over my shoulder; 'oh you have to do it now, this is what is important, you should do this or that…' No, not at all. No, no. On the contrary, for geography and history we don't use working books anymore. Lesson books, but no working books. So we work in projects. (…) and I am one of the founders, one of the ones who had the idea.
>
> (Thijs, group 8, Basisschool De Witt)

Market Accountability Teachers at the two high-performing schools perceived there to be a general, underlying pressure to maintain their schools' strong academic reputations. They reported management to place a greater value on performance and test scores, and several felt an expectation to show a continuous trajectory of improvement. This, teachers found unrealistic, likening it to private sector expectations of year-on-year profit growth:

> We are like a commercial company where you are expected to have a credit growth every year, and we're also expected to increase the results every year… but we're talking about humans here, it's not like a pile of money.
>
> (Sara, group 6, Barlaeusschool)

> Sometimes I do have the idea that I'm working for a company, working for a car factory. The quality has to be impeccable, and we don't need any deviations, it has to be a linear process (of improvement), but it isn't, we're working with kids.
>
> (Finn, group 5, Basisschool De Witt)

This reputational pressure is innately tied into the schools' academic image and predominately instrumental logics. Basisschool De Witt was not found to experience market competition, yet maintaining its reputation as one of the best schools in the city was a pressure in itself:

> In *this* class, a lot of people don't go to HAVO. A lot of children need extra instruction and they go to VMBO-T. And that is average in Holland - it's okay. But for this area here, we are 540 on Cito end-test, 60% goes to VWO... Those are our goals.
>
> (Thijs, group 8, Basisschool De Witt)

Reputation was perhaps even more pressing to Barlaeusschool, given its struggle to attract students. This market pressure was clearly felt by the vice-principal and did not go unnoticed by teachers: it was the only school where a number of enrolment strategies were highlighted. One of these strategies, as noted, was a reorientation of school practices to spend more time on non-core learning areas. According to teachers, this approach had not (yet) resulted in a reduction of performance pressure. Rather, it seems that letting go of an outcome-based raison d'être was proving difficult for the school. Test scores, quite possibly, have become integral to how Barlaeus views its own success, an image that is difficult to relinquish.

> So you have the head, the hands and the heart and we always say we try to focus on everything, but in the end, it's a lot of head.
>
> (Marianne, group 3, Barlaeusschool)

This in turn, appeared to affect Barlaeus teachers' experience of autonomy, who reported to be limited by a rigid curriculum and strict working methods, particularly when it came to the core learning areas. Respondents expressed a desire to break free from the methods and to have more fun and engagement with their students. 'I would try and have more meaningful moments' (Marianne, group 3, Barlaeusschool). Unlike at Erasmus and De Witt, teachers did not mention their involvement in the selection of these methods.

Participative Accountability In the Dutch system, standardised tests not only have stakes for schools but for students too, functioning as a key determinant of their future career path. This adds to the burden on teachers. Pressure that was described as 'self-imposed' was reported as a significant pressure source in all four schools. Not only did some teachers measure their own abilities through test results, but almost all felt a keen responsibility not to let their students down by allowing a sub-optimal performance. Even greater than this however, was the pressure teachers felt from parents. Reportedly, parental expectations were not always realistic:

> Most of the parents say… 'I want HAVO, because HAVO is the best.' Yes, HAVO is the best for *some* children, but not for your child…
>
> (Mark, group 7, Erasmusschool)

Parental pressure was present in all schools, but was reported to be a greater concern in the high performing Barlaeus and De Witt. Interviews at Barlaeusschool in particular revealed the considerable stress felt during the period when students' secondary school advice is given. It was not uncommon that parents did not see eye-to-eye with the teacher on this: 'I got into *so* many fights with really abusive parents' (Sara, group 6, Barlaeusschool).

Teachers at Basisschool De Witt did not feel this parental pressure as directly (or at least seemed to internalize it less), yet expressed significant concern about its effect on students. They understood it to be a result of the school's population; high-earning, well-educated parents with high expectations and considerable involvement in their child's education (see also Mathison & Freeman 2006). Interviews frequently revealed that many parents at the school sign-up their children for (extra-curricular) test training. This was not mentioned in any of the other schools. There were also accounts of a small number of parents one year buying copies of an upcoming test that had been 'leaked'. As well as worrying that this undermined the validity and use of the tests, teachers were significantly concerned about the stress this placed their students under. This concern was one of the key reasons why teachers in high-performing schools reported to only discuss and practice test questions in the few weeks before the test.

DISCUSSION

The aim of this study has been to better understand the impact of test-based accountability policies by examining the experiences of teachers in differently performing primary schools in the Netherlands. More specifically, it has inquired into the value that schools and teachers attach to academic performance (and standardised testing in particular), the extent to which teachers experience performance pressure, and the impacts of this on their work. Comparing two ‘low’ and two ‘high’ performing schools, the research addresses the question: In such performative environments, (how) does school performance level affect teacher experience?

Findings show that schools operating within the same TBA structures under the same management are confronted with quite different challenges. It is clear that accountability mechanisms do not have an equal impact, rather, the four schools were found to be vulnerable to different (combinations of) pressures. The lower performing schools are, by policy design, under greater external administrative pressures than their high performing counterparts. Echoing the experiences of probation schools in the US (Diamond & Spillane 2004; Mintrop 2004), Van Schurmanschool was found to be experiencing the most immediate challenges and direct pressures due to the school-level interventions attached to underperformance, and in response was instigating various, and reportedly superficial, educational changes. Market pressure was not found to be related to school performance; the ‘underperforming’ Van Schurmanschool and the ‘high performing’ De Witt had waiting lists, whilst Barlaeusschool, also high performing, was under considerable pressure to attract new students. This has more to do with the schools’ logics of action. Whereas the other schools were currently occupying a particular position within the education market (Maroy & Van Zanten 2009), Barlaeus’ position had become jeopardised. The school was attempting to adjust its educational offer as a result.

Indeed, we see considerable interschool variances regarding the ways in which the schools manoeuvre themselves in response to accountability pressures, and the significance placed on performance. ‘Instrumental’ and ‘expressive’ are, of course, ideal types, yet clear consistencies or ‘logics’ could be found, unique to each school (Ball & Maroy 2009). Van Schurman and Erasmusschool were consistent in their emphasis on student needs, equality and diversity. This was also apparent in their organisation of education and approach to testing. In comparison, Barlaeus and Basisschool De Witt were much more explicit in the promotion of their results and perceived good performance as central to their reputation. There were further differences regarding the expected role of teachers and relational structures in the schools. Collaboration and an ‘open-door’ approach were repeatedly emphasized as central to Basisschool De Witt’s culture, yet not mentioned at Barlaeus or Van Schurman. Similarly, meaningful teacher involvement in *core* curricular decisions was only reported at Erasmus. As will be discussed shortly, the importance and impact of these logics of action on teachers appears significant, apparently explaining the unexpectedly low or high levels of pressure they feel, regardless of school performance.

Despite differing external pressures and internal logics, the considerable role of standardised testing in shaping education is evident across the schools. All four conduct standardised tests from an early age, follow prescribed methods, engage in test preparation activities, practice ability differentiation on the basis of results, shape staff meetings, school reports and teacher monitoring and evaluation around results, and keep careful records of student learning. In this way, test data used as a representation of learning and signifier of quality has become entrenched in Dutch schools as it has in high-stakes neoliberal systems (Ball 2016; Holloway & Brass 2018). Essentially, regardless of the 'autonomy' label so often afforded to Dutch schools, performance pressure, stemming from a variety of sources, restricts this freedom.

Inevitably, these pressures have, in turn, impacted teachers' work and their own experiences of performativity and professionalism, and reflect what is best understood as a market model approach to professional regulation (Voisin and Dumay 2020). When considered alongside prominent studies conducted in other 'market model' contexts, (Day & Smethen 2009; Perryman et al 2011; Valli & Buese 2007), they indicate a commonality to teacher experience that stands apart from the accountability structures in place. Indeed, in the Netherlands no formal administrative or economic stakes are attached to 'teacher performance', yet various other mechanisms have closely tied this performance to reputation, professional identity, and students' futures. In this way, TBA policies have changed the very identity of teaching; both from outside and inside the profession (Ball 2016).

While some elements of TBA have become naturalised or taken for granted, frustration and dissatisfaction amongst teachers remained apparent. Prescribed and test-heavy working methods may be considered the first line offender here, yet they do not work alone. Time and scheduling pressures, exacerbated by the burden of accountability-related admin means that even teachers who described themselves as autonomous, confessed to feeling restricted by their workload. Interschool variances were evident in teachers' experiences of their work and performance pressure. School performance and logics of action play an important role here. Although TBA interventions in the Netherlands are designed to have a primarily informative and supportive role – contrasting with the higher-stakes sanctioning role described by Mintrop (2003) in the US – their impact on teachers is nonetheless significant. Despite Van Schurman's primarily expressive orientation, teachers perceived there to be a substantial, even overwhelming, focus on test results. These teachers reported considerable performance pressure and restricted autonomy. In contrast, despite also finding itself at the low end of the performance spectrum, teachers at Erasmus spoke of their work satisfaction and claimed to experience relatively low results-based pressure. Further, rather than finding the system unfair, the principal welcomed the comparison of her school with others. Two key differences exist here. Firstly, Erasmusschool has managed to achieve consistent satisfactory results, meaning it is not subjected in the same way to external administrative pressures as Van Schurman. Secondly, while the two schools share many of the same priorities and 'expressive' orders, unlike Van Schurman, Erasmus seems to also adopt a collaborative

decision-making approach, with teachers reporting a meaningful involvement in key curricular decisions. This is significant: despite being one of the schools most focused on core learning, Erasmus' teachers expressed the greatest sense of control over their work. Indeed, even in 'disadvantaged' schools where educational content may be narrowed in an effort to meet core quality standards, teachers may still be able to enjoy a meaningful sense of curricular and pedagogical freedom.

A more expressive approach to educational decision-making was also reported at De Witt. Teachers' involvement in (non-core) curricular choice appears to provide them with a deeper understanding and greater sense of ownership over what they are teaching, and with a greater chance of working with methods they feel are suited to their students. As such, policy is experienced more as something they are *doing*, rather than as something being *done to* them (see Ball et al 2011). Despite recognising the presence of external pressures, teachers at Erasmus and De Witt believed their schools to have clear and collective missions that were not susceptible to changeable government policies and societal expectations.

Despite the 'autonomy' of the Dutch school system therefore, and despite what findings from studies focussing on *administrative* pressure may suggest, this research indicates that in a system so shaped by accountability, it is inaccurate to assume that well-performing schools and their teachers will experience no or low performative pressure. It is true that administrative sanctions or 'actions' should not trouble such schools, but pressures come from elsewhere. In the more instrumentally-oriented Basisschool De Witt and Barlaeusschool, where image and market position are so dependent on performance, teachers reported to feel greater pressure from management and parents: the pressure to 'stay on top.' In De Witt, where the school and its board applied for a quality label, the importance of reputation also resulted in the stress and increased workload associated with inspection. Similarly, it is important to emphasise that low performing schools and their teachers will not necessarily experience high degrees of performance pressure. Due to adequate performance, a primarily 'needs over outcomes' focus, and an inclusive management approach, teachers at Erasmus reported to experience some of the lowest levels of pressure across the four schools.

CONCLUSIONS

This study has examined the enactment and impact of TBA at high and low performing primary schools in what can be described as a 'middle-stakes' system. It has shown that, despite the differences that exist between the Dutch and those 'higher-stakes' Anglo-Saxon systems explored in the paper, notable similarities in policy impacts are apparent. An interplay of various accountability mechanisms create pressure on different parts of the education system which is significant in steering schools' and, inevitably, teachers' practices and experiences. As a result, perceived performance pressures within the

school may be considerable, even when administrative accountability pressures are not. Conversely, policy tools designed to exert pressure on schools may not always achieve their intended impact.

Despite adopting many of the same test-based practices, schools place different significance on performance. The nature and sources of performative pressures vary between high and low performing schools, yet – given 'sufficient' performance – it cannot be said that one context is *less* pressured than the other. While teachers in the more 'disadvantaged' schools felt their work to be, in many ways, more challenging and more conditioned by TBA, findings suggest that a more privileged student population is no guarantee of escape. Rather, schools and their teachers are susceptible to performative pressures in various ways and for various reasons. Overall, teachers expressed higher levels of professional satisfaction and autonomy in the schools with a clear and collective mission, reporting to be less vulnerable to changing government demands and external pressures. Further research, exploring this relationship and its direction is warranted here.

While this study does not purport to make broad generalizations, it does reveal the various pressures – shared and diverging – felt by teachers in schools that face different challenges within the context of performativity. In doing so it adds to the growing body of research that draws attention to the assumptions and oversights on which these, now globally-spread, accountability policies are based.

REFERENCES

Agrey, L. (2004). The Pressure Cooker in Education: Standardized Assessment and High Stakes. Canadian Social Studies, 38(3), 1-12.

Ball, S. (2003) 'The teacher's soul and the terrors of performativity' Journal of Education Policy, 18 (2), 215-228

Ball, S. & Maroy, C. (2009) School's logics of action as mediation and compromise between internal dynamics and external constraints and pressures, Compare: A Journal of Comparative and International Education, 39:1, 99-112,

Ball, S. J. (2016). Neoliberal education? Confronting the slouching beast. Policy Futures in Education, 14(8), 1046-1059.

Ball, S., Maguire, M., Braun, A., & Hoskins, K. (2011). Policy actors: Doing policy work in schools. Discourse: Studies in the Cultural Politics of Education. 32. 625-639

Bernstein, B. (1996). Pedagogy symbolic control and identity theory, research, critique (London & Bristol, Taylor and Francis).

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.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework 1. *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

Browes, N., & K Altinyelken, H. (2019). The instrumentation of test-based accountability in the autonomous Dutch system. *Journal of Education Policy*, 1-22.

CNV Onderwijs & EenVandaag (2018). *Onderzoek: 'Leraren onder druk om schooladvies.'* Available at:https://eenvandaag.avrotros.nl/fileadmin/user_upload/PDF/Rapportage_Peiling_Leraren_Schooladvies.pdf

Collier, D., & Mahoney, J. (1996). Insights and Pitfalls: Selection Bias in Qualitative Research. World Politics, 49(1), 56-91

Dale, R. (1982), 'Education and the Capitalist State: Contributions and Contradictions', in Apple, M (ed.), Cultural and Economic Reproduction in Education, London: Routledge and Kegan Paul

Darling-Hammond, L. (1990) Teacher professionalism: why and how, in: A. Lieberman (Ed.) Schools as collaborative cultures: creating the future now (London, Falmer Press).

Day, C. (2002). "School reform and transitions in teacher professionalism and identity". *International journal of educational research*, *37*(8), 677-692.

Day, C., & Smethem, L. (2009). The effects of reform: Have teachers really lost their sense of professionalism? Journal of Educational Change, 10(2), 141-157.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality?. *Teachers college record*, *106*(6), 1145-1176.

Dobbins, K. (2009). Teacher creativity within the current education system: a case study of the perceptions of primary teachers. Education 3-13 37 (2):95-104

DUO Onderwijsonderzoek & De Monitor (KRO-NCRV). (2016). Rapportage: Werkdruk leerkrachten in het basisonderwijs.

Hargreaves, D., (1995) School Culture, School Effectiveness and School Improvement, School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 6:1, 23-46

Holloway, J. & Brass, J. (2018) Making accountable teachers: the terrors and pleasures of performativity, Journal of Education Policy, 33:3, 361-382

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277-1288.

Inspectie van het Onderwijs. (2020) Onderzoekskader 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs.

Inspectorate of Education. (2019). State of Education, 2020, Summary Report. Utrecht: Inspectie van het Onderwijs.

Maroy, C. & Van Zanten, A. (2009). Regulation and competition among schools in six European localities. *Sociologie du travail*, *51*, e67-e79.

Mathison, S., & Freeman, M. (2006). Teacher stress and high stakes testing. *Understanding teacher stress in an age of accountability*, *Greenwich, Conn: IAP.* 43–63.

Mintrop, H. (2003). The Limits of Sanctions in Low-Performing Schools. *Education policy analysis archives* 11:3

Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. Teachers College Press.

Mintrop, H. (2007). School Improvement Under Test-Driven Accountability : A Comparison of High- and Low-Performing Middle Schools in California CSE Report 717 Heinrich Mintrop National Center for Research on Evaluation , Standards , and Student Testing (CRESST) / Universit, 1522(310).

Nathaniel, P., Pendergast, L. L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states. *Teaching and Teacher Education*, *59*, 492-502.

Nusche *et al.* (2014). “School education in the Netherlands”, in *OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014*, OECD Publishing, Paris.

OECD (2010), PISA 2009 Results: What Makes a School Successful? – Resources, Policies and Practices (Volume IV), PISA, OECD Publishing

OECD (2013), PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV), PISA, OECD Publishing

Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods: Integrating Theory and Practice* (4th edition). Thousand Oaks, CA: Sage.

Perryman, J., Ball, S., 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

Saha, L. J., & Dworkin, A. G. (2009). *Teachers and Teaching in an Era of Heightened School Accountability: A Forward Look*. In *International handbook of research on teachers and teaching* (pp. 1177-1185). Springer, Boston, MA.

Thiel, C. & 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)

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, *44*(3), 519-558

Van Zanten, A. (2009) Competitive arenas and schools' logics of action: a European comparison, Compare: A Journal of Comparative and International Education, 39:1, 85-98

Verger, A., & Parcerisa, L. (2017). A Difficult Relationship: Accountability Policies and Teachers. International Evidence and Key Premises for Future Research. In M. Akiba & G. LeTendre (eds.). International Handbook of Teacher Quality and Policy (pp.241-254). New York: Routledge.

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

Voisin, A., & Dumay, X. (2020). How do educational systems regulate the teaching profession and teachers' work? A typological approach to institutional foundations and models of regulation. *Teaching and Teacher Education*, *96*, 103144.

Waslander, S. (2010). Government, School Autonomy, and Legitimacy: Why the Dutch Government is Adopting an Unprecedented Level of Interference with Independent Schools. *Journal of School Choice*, *4*(4), 398–417.

Webb, R., Vulliamy, G., Hämäläinen, S., Sarja, A., Kimonen, E., & Nevalainen, R (2004): A comparative analysis of primary teacher professionalism in England and Finland, Comparative Education, 40:1, 83-107

West, A., Mattei, P., & Roberts, J. (2011) Accountability and Sanctions in English Schools, British Journal of Educational Studies, 59:1, 41-62,

Wilkins, C. (2011). Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system. *Professional development in education*, *37*(3), 389-409.

Yin, R.K. (2009), Case Study Research: Design and Methods, 4th ed., e-book, Sage, CA (accessed 20 December 2010).

CHAPTER 6

GLOBAL SUMMARY OF FINDINGS

The aim of this research has been to answer the question:

> ***How and why have school autonomy and test-based accountability policies in the Netherlands evolved over recent years, and what impact has this had on teachers' practices, beliefs, and overall professional experiences?***

To do so it has focused on both policy making and policy enactment, and has looked to meta, macro, meso, and micro level factors and the relationships between them. Below, in answer to this question, the main findings of the study are presented, organised by each sub-question theme.

The Evolution of School Accountability in the Netherlands & the Role of School Autonomy

The development and evolution of TBA in the autonomous Dutch education system has been found to be a process influenced by a complex interplay of multi-level factors. A cultural-historical perspective reveals that many of the cornerstones of SAWA were present in the system before NPM reforms started to take hold in the 1980s, with some existing for over a century. However, these policy instruments have been enhanced and recalibrated over recent years to more closely resemble the SAWA governance system that has become globalised, with TBA developing incrementally and testing taking an increasingly central role in quality assurance over a period of several years.

Significantly, the reasons for the adoption of TBA policies were found to differ from the Anglo-Saxon early-adopters. That is, rather than intended to stimulate school choice and market dynamics (see Verger, Parcerisa & Fontdevila 2019), TBA in the Netherlands has primarily been justified through quality concerns, and seeking an autonomy-accountability balance. Increased decentralisation which shifted key managerial responsibilities to the schoolboard and school levels resulted in the problematisation of excessive and unchecked school(board) autonomy and slipping system quality. The implementation of learning standards and compulsory standardised testing followed these widespread, yet somewhat unsubstantiated concerns.

TBA policy tools were not only chosen for their perceived effectiveness but also for their legitimacy, or suitability to the Dutch context (Kassim & Le Gales 2010). Freedom of education played a significant role in calibrating these tools, and in many ways, acted to soften them. At the primary level, testing

became compulsory and used for school accountability purposes but, after some negotiation, it also became more flexible, allowing, for example, schools to choose their test provider. This in turn has had considerable impacts on the testing market and raised issues of test comparability. These findings point to the necessity of a having a contextualised understanding of policy adoption that respects the key role of the 'translation' process (Steiner-Khamsi 2014) in recalibrating global policies. Macro level perspectives miss this. The, at times unexpected, journey of TBA in the Dutch system also supports the notion of policy tools as life forms, changing independently of those decisions that created them and producing unintended consequences (Lascoumes & Le Gales 2007).

The Role of International Organisations in the Promotion of SAWA &TBA

Data has indicated that a prominent factor within this policy reform has been the exposure to, and engagement with, international ideas. Interview accounts and policy documents demonstrate an awareness amongst Dutch policy experts of the internationalisation of particular policy practices and their promotion by transnational stakeholders, particularly the OECD. In the early 2000's, international comparative data, particularly PISA performance, contributed to the construction of policy problems and compliance pressures. However, particular externally-promoted ideas were also found to be used strategically by the Dutch government as a way of legitimising reform, known as 'externalisation' (Steiner-Khamsi 2004). To this end, as noted by Savage and Lewis (2019 p.134): 'We see the OECD's ideas and agendas being taken up and re/dis/assembled into specific national contexts.' Indeed, accountability-based reform certainly suited the government's political agenda of the early 2000's.

Findings have also highlighted that the international SAWA agenda is promoted through other means. Specifically, the role of the OECD's policy peer reviews has been examined. The influence exercised through such reviews exists in often subtle ways, but is nonetheless a powerful tool with which to shape education governance on a global scale. This includes the prominence of SAWA policy recommendations and the problematisation of 'misaligned' policies throughout the review, as well as the regular use of competition and benchmarking pressures. Beyond this, it is important to highlight the influence exercised through the socialization and negotiation processes that take place when engaging in peer reviews (Grek 2017). In these ways, the reviews contribute significantly to the 'ideas game': 'We see the OECD playing a particularly powerful role informing what concepts like 'student achievement', 'teacher quality' and 'effectiveness' mean and look like, both in policy and practice' (Savage and Lewis 2018 p.136). The Dutch government is inducted into the 'rules of the community' and subtly pressured to adopt SAWA practices as the 'right thing to do' (Grek 2017). Significantly, being only loosely coupled to specific policy issues, autonomy and accountability solutions can be attached to a vast array of problems. This provides great flexibility, a characteristic that has undoubtedly facilitated SAWA's global spread.

By conducting such reviews, the OECD's purpose and status in educational expertise is cemented (see for example, Grek 2016), yet this is only one side of the story. Departing from globalist views that overestimate the influence of globalisation and IOs (see Martell 2007), attention has also been paid to the reasons for participation on the Dutch side. OECD recommendations were viewed by the MoE as apolitical and scientifically based, and as a way to redirect issues that had become too sensitive at the national level into more 'technical' and desirable channels. Rather than intentionally used to legitimate specific reforms or policies therefore, the review was intended to help reframe the national education debate. Yet, the review did not deliver the insights the MoE had hoped for, and any immediate or direct impact on policy was limited. Given that thus far, studies in this field have focused on those contexts in which reviews had considerable policy impact (see Grek's 2017 analysis of the Swedish review), this finding is significant, and highlights the importance of the reviewer-reviewee relationship. As a 'good student' the OECD was seen to take a highly cooperative, non-patronising role throughout the process. Further, unlike in the Swedish context referenced, in the Netherlands, the OECD *contributes to* but *does not lead* the education debate. Nonetheless, the review still had important political uses. The MoE for example, was able to make selective use of recommendations to justify particular reforms and help quash growing, undesirable sentiments against testing.

The Impact of TBA & Performative Work Environments on Teachers' Practices, Beliefs & Professional Identities

Testing, it has been found, has come to take a central position in Dutch primary schools, at least those involved in this study, and the impact of TBA on teachers appears significant, affecting their everyday tasks and notions of what it means to be a good student and a good teacher. While testing itself is not viewed negatively by teachers, many appeared frustrated with the importance it has developed and the wider repercussions of this. While, as argued, TBA in the Netherlands has developed in a softer way than it has in England, and in US contexts which dominate such research, it is striking that teachers in the Dutch system experience many of the same undesirable effects on their work.

Teachers reported a high task load, much of which includes results-driven administrative work; developing and reviewing class plans, forming differentiated student groups, and closely recording the work of these groups to ensure students stay on track with the core standards and testing schedules. Teachers commonly felt overburdened with tasks and struggled under time pressures. As seen time and again in high-stakes contexts, this has served to constrict classroom practices in terms of what, how, and at what pace students are taught (see Berryhill, Linney & Fromewick 2009). As found by Valli and Buese in 'one mid-Atlantic' state, role change (including role increase) has resulted in teachers in the Netherlands reporting to have insufficient time to plan interesting, thought provoking lessons and to ensure students have full understanding of a topic before moving onto another (Valli & Buese 2007). Adding to this, it has also found that teachers often have to skip tasks, in particular checking and marking

students' classwork and homework. Indeed, the time required to constantly monitor the learning process seems to have markedly reduced time for learning itself. The similarities between the Dutch and the US contexts referenced above, corroborates quantitative studies such as that by Thiel and Bellmann (2017): it cannot be assumed that undesirable TBA effects are attached only to high-stakes contexts.

Findings do not only reveal clear inter-contextual similarities, but inter-professional ones too. Drawing parallels with quantitative studies conducted amongst nurses (Takase, Maude & Manias 2006), school social workers (Agresta 2006), and occupational therapists (Lloyd, King & McKenna 2004), role discrepancy amongst teachers was found to be commonplace. Like these professionals, the majority of teachers wanted to spend more time on their areas of specialism (teaching and planning lessons) and less time on generic, administrative tasks (see Agresta 2006; Lloyd et al 2004). Yet moving beyond these general trends, the qualitative nature of this study was also able to uncover the different ways that teachers perceived and responded to role discrepancy. Some felt forced to choose between job requirements and their own professional beliefs about what it means to be a good teacher. Others accommodated both required and desired work by fitting performative tasks into their usual schedule, or by making out-of-hours work a part of their working week. A minority of teachers adopted a mixture of these positions, as deemed practical. These findings complement those conducted within Dutch secondary schools, which uncovered similar teacher responses, albeit through a different conceptual approach (Hendrikx 2019). Yet they also move beyond this, identifying more nuanced teacher responses that reflect important variations in professional beliefs, and the significance of professional adaptability. Indeed, not all participants could be understood to be experiencing role discrepancy; a small minority of teachers, all early in their career stage, were able to align their professional beliefs with the demands of TBA, seeing performative tasks as contributing to their students' achievement within the existing educational framework.

These findings offer new ways of understanding teacher experience of, and response to, accountability policies. They support research which suggests NPM reforms have, and continue to, reshape the teaching profession; not only what is expected of teachers, but what teachers expect for themselves and from their work (Ball 2016). A significant number of participants spoke about fundamental aspects of the profession – autonomy, collegiality and trust – as luxuries, and as something to be earned. Yet others, predominately those in the profession over a decade, held on to a more 'traditional' notion of professionalism. This latter group felt frustrated by the perceived erosion of what they saw as core aspects of the profession, and were found to be struggling to adapt to the new realities of teaching that policy reform has brought. The extent to which this can be attributed to prior policy exposure, as argued by Holloway and Brass (2018), and the extent to which it might (also) be attributable to career-cycle (Hargreaves 2005) is not clear. Yet, there is certainly an indication that the teaching profession is being reshaped by TBA and the data-driven and performative demands that accompany it. Despite a small sample, the data suggests a divergence of professional identities (Noordegraaf 2016), with transitioning

notions of the ‘good teacher.’ This is witnessed in the divide between those teachers who are overall content with their work, the ways in which they are managed, and how their abilities are judged, and those who feel professionally unfulfilled, seeking greater trust in their experience and expertise.

Teacher Experience of TBA & Performance Pressure in Low & High Performing School Contexts

Finally, given the unequal and unintentional impacts that policy tools have been proven to have (see chapter 2), it was considered important to examine the impact of contrasting school contexts on teachers' experiences of TBA. Analysing two high and two low performing schools, it seems that testing and test data have become a pivotal part of school life in both contexts – understood as a core way of measuring teaching, learning, and overall education quality. This reflects that TBA has planted strong roots in Dutch schools as it has in high-stakes contexts, resulting in teachers that are under pressure to ‘perform;’ to improve or maintain results and show proof of results-oriented working. Yet inter-school comparisons also reveal that different schools are confronted with different challenges. Perhaps unsurprisingly, principals and teachers in the low-performing schools felt under greater strain from administrative accountability measures. Similar to studies conducted in the US (see for example Achinstein & Ogawa 2006; Mintrop 2003), these schools had high proportions of students for whom Dutch was not their native language, and teachers perceived an innate sense of unfairness to accountability – of being held to the same standards as those more ‘privileged’ schools. These pressures resulted in a clear focus on the core learning areas, and a frustration that students were missing out on broader aspects of education. Performance and administrative-based pressures were found to be most prevalent within the school deemed ‘underperforming’ and subject to ongoing inspection. This school's status is comparable to probation schools such as those within the contexts of Mintrop's (2004) and Diamond and Spillane's (2004) US studies. Like these schools, the primary goal of this underperforming Dutch school was a change of quality status, and teachers felt pressured to incorporate changeable, superficial education approaches to achieve this goal.

Yet, pressure does not only come from administrative sources. Reputational and market pressures were also found to play a significant role. Regarding the former, the importance schools placed on results is tied into their ‘logics of action’ (see Ball & Maroy 2009): teachers in the high performing schools, built on strong academic reputations, were found to experience pressure from parents and from management to ensure this reputation was realised and results maintained. One ‘high-performing’ school had also recently undergone an (elected) inspection in the hope of obtaining a quality label. While, as noted by Diamond and Spillane (2004), this reflects that responses in well-performing schools are more likely triggered by rewards rather than sanctions (the lure of a ‘good school’ label rather than the threat of a ‘very weak’ one), it nonetheless placed teachers under many of the same administrative and workload strains as those in the ‘underperforming’ school. For the other high-performing school in the study,

market competition was perceived as a considerable pressure source. Despite its good record of performance, the school was struggling to attract sufficient student numbers to maintain funding. Indeed, parents' decision-making around school choice is not limited to performance, but is considerably more complex (see Maroy & Van Zanten 2009).

As well as the emphasis placed on academic achievement, a school's pedagogical and managerial approach, another core aspect of its logics of action (Ball & Maroy 2009), appears to play an important role in teachers' experiences of pressure and their work. Teachers who felt meaningfully involved in curricular decisions reported a lesser degree of frustration with TBA demands, regardless of school performance level and educational approaches. These teachers felt more supported and trusted by their management and reported higher levels of professional autonomy. In comparison, in schools where such decisions were perceived as more top-down, teachers felt less in control of their classrooms and frustrated by requirements to follow curricula and methods they did not always agree with. This implies that, with the right management, even in schools under pressure to narrow educational content in order to achieve minimum required results, teachers may still be able to enjoy a meaningful sense of control and freedom over their work.

REFERENCES

Achinstein, B., & Ogawa, R. T. (2006). (In)fidelity: what the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, *76*(1), 30–63.

Agresta, J. (2006). Job Satisfaction Among School Social Workers. *Journal of Social Service Research*, *33*(1), 15–26.

Ball, S. & Maroy, C. (2009) School's logics of action as mediation and compromise between internal dynamics and external constraints and pressures, Compare: A Journal of Comparative and International Education, 39:1, 99-112.

Ball, S. J. (2016). Neoliberal education? Confronting the slouching beast. *Policy Futures in Education*, *14*(8).

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.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality?.*Teachers college record*, *106*(6), 1145-1176.

Evetts, J. (2009). New Professionalism and New Public Management: Changes, Continuities and Consequences. *Comparative Sociology*, *8*(2), 247–266.

Grek, S. (2016). Beyond the standardisation vs. contextualisation debate: The role of the OECD in European education governance. Paper presented at the International Organizations and the Globalization of Public Instruments and Ideas (IOGLOB) Seminar, Strasbourg.

Grek, S. (2017). Socialisation, learning and the OECD's Reviews of National Policies for Education: the case of Sweden, Critical Studies in Education, 58:3, 295-310

Hargreaves, A. (2005). Educational change takes ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education*, *21*(8), 967–983.

Hendrikx, W. (2019). What we should do Vs what we do: Teachers' professional identity in a context of managerial reform. Educational Studies, 1-17

Holloway, J. & Brass, J. (2018) Making accountable teachers: the terrors and pleasures of performativity, Journal of Education Policy, 33:3, 361-382

Kassim, H., & Le Galès, P. (2010). Exploring governance in a multi-level polity: A policy instruments approach. *West European Politics*, *33*(1), 1–2

Lascoumes, P., & Le Gales, P. (2007). Understanding Public Policy Through its Instruments: From the Nature of Instruments to the Sociology of Public Policy. *Governance*, *20*(1), 1–21

Lloyd, C., King, R., & McKenna, K. (2004). Actual and Preferred Work Activities of Mental Health Occupational Therapists: Congruence or Discrepancy? *British Journal of Occupational Therapy*, *67*(4), 167–175.

Maroy, C., & Van Zanten, A. (2009). Regulation and competition among schools in six European localities. *Sociologie du travail*, *51*, e67-e79.

Martell, L. (2007). The third wave in globalization theory. *International studies review*, *9*(2), 173-196.

Mintrop, H. (2003). The limits of sanctions in low-performing schools. *Education policy analysis archives*, *11*, 3.

Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. Teachers College Press.

Noordegraaf, M. (2016). Reconfiguring professional work: Changing forms of professionalism in public services. *Administration & Society*, *48*(7), 783-810

Porter, T., & Webb, M. (2007). The Role of the OECD in the Orchestration of Global Knowledge Networks (Papers prepared for presentation at Canadian Political Science Association annual meetings Saskatoon, Saskatchewan, Canada, Wednesday May 30, 2007).

Savage, G. C., & Lewis, S. (2018). The phantom national? Assembling national teaching standards in Australia's federal system. *Journal of Education Policy*, *33*(1), 118-142.

Schuller, T. (2005). Constructing international policy research: The role of CERI/OECD. European Educational Research Journal, 4(3), 170–180.

Steiner-Khamsi, G. (Ed.). (2004). The global politics of educational borrowing and lending. New York: Teachers College Press.

Steiner-Khamsi, G. (2014). Cross-national policy borrowing: Understanding reception and translation. *Asia Pacific Journal of Education*, *34*(2), 153-167.

Thiel, C., & 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).

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, *44*(3), 519-558.

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

CHAPTER 7

CONCLUSIONS

DISCUSSION OF MAIN FINDINGS

Based on the key findings, presented in the previous chapter, a number of points are worth highlighting, both in terms of what they reveal about the impacts of SAWA in the Netherlands, and because of what they reveal about the nature and implications of these reforms more broadly.

Firstly, the effects of SAWA revealed in this study appear to contradict some of the theories and assumptions on which the model is based. Data indicates that the traditionally diverse and autonomous Dutch education system might not have been enhanced by SAWA, but rather, impaired by it. This is perhaps the result of contrary mechanisms within the governance model itself; reaching diversity through standardisation and autonomy through accountability seem like challenging goals. Rather, narrow measurements of quality, unclear government requirements, and an education market crowded with stakeholders has resulted in Dutch schools following prescribed and restrictive methods that are formulated around test content. This leads to pedagogical and curricular convergence, and restricts teacher autonomy. These contradictions are not only true of administrative, but also of market mechanisms. Publishing results does not appear to have shaped school choice and stimulated school response as expected. Indeed, in the small number of schools analysed, this study has not found school performance to be related to student enrolment numbers. Concern over performance indictors has been seen to be more related to school culture than school performance, with the high-performing schools in the study seemingly under the most pressure to maintain their academic reputation amongst parents. Relying on market mechanisms to improve low performance does not therefore appear realistic. Further, both the administrative and market tools are based on the (ill-founded) assumption that schools and teachers have a critical role in shaping student outcomes.

Secondly, this study highlights the importance of conducting more research in systems where SAWA and accountability tools have been adopted for different reasons and developed in different ways than in those heavily-researched, early adopters of the reform. In contexts such as these, where accountability appears softer and does not target teachers directly, policy-enactment studies risk being neglected, and policy impacts ignored or diminished. Yet, as discussed in chapter 6, the similarities found between this study and comparable studies conducted in the US and the UK, is striking. Work and performance

pressures are very much a part of Dutch teachers' everyday lives: many feel overburdened with tasks, particularly administrative ones, and that they are lacking the time and the freedom to explore and enjoy (new) topics. This points to the significance of professional and reputational pressures. Teachers are not striving for good student results because these results are connected to their salary; they are striving for good results because test scores have become inherently tied into the notion of teacher and school quality, because results are crucial to their students' futures, and, due to these factors, because they experience pressure from management and from parents to perform.

Thirdly, adopting a broader perspective, it has been shown that the experiences of teachers are not unique, but are in many ways reflective of those of a large number of professionals working within environments shaped by new public management (see for example Hendrikx & Gestel 2017). Reforms have brought with them privatisation, performance targets, and surveillance, and are altering the concept of professionalism. Not only have *professions* been shaped from the outside, but *professionals* have been shaped from within; adjusting their values, expectations and very identities (Ball 2016; Noodegraaf 2016). This is perhaps evident in the differences reported in this study, between how teachers perceive their work: those who are generally accepting of work tasks and the role of testing and data, and those who crave greater autonomy, fewer managerial demands, and who value process rather than results. While further research is necessary here, the data suggests that career stage may be central to these differences. As time marches on, the number of teachers who have only worked within these policy environments increases, teacher training becomes more closely aligned with policy demands, and teachers enter the profession who have themselves only been schooled within SAWA environments (see Wilkins 2011). As such, the concept of what it means to be a teaching professional will, it seems, continue to change.

Fourth and finally, it is worthwhile connecting the findings on the nature of policy development (discussed in chapters 2 & 3) with the findings on policy enactment and impact (chapters 4 & 5). Doing so shines light on the complications that can result from the complex and often fragmented nature of policy-making. TBA was not found to be assembled as a coherent 'whole' but in a piecemeal manner, influenced by various (changing) factors at multiple levels. This is reflective of the concept of 'policy assemblage' (see Savage & Lewis 2018). It has resulted in accumulating demands on practitioners, and new policy tools that undermine existing ones. The teachers interviewed in this study reported to have a considerable number of demands upon them, yet they often found the source and the purpose of these demands unclear. Messy, piecemeal policy-making, influenced by a concoction of (sub) national and transnational ideas, has driven the increasing importance of testing and data in schools, and overtime, appears to have resulted in layers of unnecessary requirements on teachers.

LIMITATIONS & FURTHER RESEARCH

This study has relied on in-depth, qualitative data and document analysis, to uncover and explain the causal mechanisms behind the development of SAWA education policies and the enactment of these policies. Importantly, this rich data has enabled new insights into the impacts of policy on the practices and beliefs of teachers in the Netherlands. One way this data could have been made richer still, is through the use of classroom observation. Such observations could provide the researcher with a greater understanding of the contexts in which teachers work and the challenges they face. Further, triangulating observations with interviews accounts can help limit the occurrence of (unintentionally) false or socially-desirable reporting (Hammersley & Gomm 2008). Yet, given that the research was primarily concerned with teachers' own beliefs and perceptions of their work, this concern was minimal. Further, due to the number of schools and teachers involved in the research, time to make comprehensive observations would have required extra researchers, or a deeper but narrower empirical strategy that would not have adequately addressed the study's research questions. Unfortunately, due to teachers' heavy workloads and busy classrooms, principals were also not always in favour of extended researcher presence.

In parallel to the qualitative data underpinning this study, quantitative research tools have also been developed. These will enable key findings from this thesis to be investigated more broadly, and for new research avenues to be explored. The quantitative research consists of comprehensive online surveys that have been sent to principals and teachers in primary schools across the country. The surveys enquire into SAWA practices, pressures and perceptions at the school and teacher levels. Such data will reveal broader patterns and indicate how representative this study's findings are of the Netherlands as a whole. For example, this data could indicate whether performance pressure is indeed commonplace amongst all schools, whether pressure sources vary according to performance level in the same way as found in this study, and whether there is a statistically significant relationship between teachers' perceptions of TBA and their career stage.

Outside of this study's parameters, such quantitative data could also be used to show a wealth of further information about the enactment and impact of SAWA in the Netherlands. As explained, in the Dutch system, teachers are not the direct subjects of TBA, rather, schools are. Yet performance pressure amongst teachers still prevails. Understanding how this pressure manifests further up the management pyramid by examining principals' perceptions of, and responses to TBA, could therefore be significant. For example, are teachers' own experiences primarily affected by those of their principals', or are there are other significant, individual level factors, at play, independent of school context?

Similarly, given the decentralised Dutch governance structure, what role school boards (the bodies officially responsible for school quality) play in schools' and teachers' experiences of TBA is also an important question. While this role is acknowledged, it was beyond the realms of the study to be

investigated fully. It is logical to assume that how school boards interpret and respond to government requirements will directly affect the expectations placed on their schools, impacting teachers' practices and perceptions of performance pressure. The large, professional board responsible for the four schools analysed in chapter 5 of this study was perceived by principals to adopt a non-intrusive approach to management, and, in the face of the Inspectorate's quality decision, was reported to have defended and supported its 'underperforming' school. Yet, boards who are particularly concerned about quality assurance may well prove more demanding, and place considerable attention on monitoring and improving results. It is also important to note that Dutch school boards vary substantially in terms of their size, structure, expertise and resources. A study that compares boards and the management of their school(s) in response to TBA could prove enlightening, helping us better understand the chain of accountability and performance pressure in a decentralised system.

Finally, as explained, one of the key characteristics of the Dutch education system is its diversity. Schools can be based on various pedagogical, philosophical or religions foundations. Some schools' approaches diverge more from the core principles of TBA than others, meaning they are caught between discourses of child development theories on the one hand and standards and test-based 'back to basics' policy approaches on the other (see also, Braun & Maguire 2018). A significant number of Dutch schools follow a Montessori or a Dalton approach for example, based on the notion that students learn in their own way and at their own pace. A growing number of schools are also classed as 'free schools' (Waldorf schools) – 'free' referring to autonomy from (undue) government intervention. This model of education values holistic learning and creative development. It is important to understand how such schools function in a system that holds them accountable through narrow measures of education quality and whether, for example, educational values are sacrificed to ensure that attainment targets are met.

RECOMMENDATIONS FOR POLICY & PRACTICE

Beyond its academic contribution, this thesis has attempted to shine a light on education policy for a more practical purpose. A better understanding of how policy is enacted and what impacts it has can contribute to evidence-based policy-making that addresses problems effectively whilst minimising unexpected and unwanted side-effects. As such, the following recommendations are made in a hope to make some small contribution to future education policy and practice.

Clarify central regulations: Often, policy implementation in the Netherlands does not take a direct route from government to schools. Rather, the decentralised system consists of various layers of accountability and a web of interconnected stakeholders through which policy is steered (see Waslander, Hooge & Drewes 2016). From discussions with principals and teachers, it is evident that centralised administrative regulations are not clear, rather, they have become convoluted through the interests of

these various stakeholders. This has added to the administrative burdens on teachers. Although this problem has been acknowledged, – in 2017 the Education Inspectorate released a guide for schools entitled 'Room in rules' (Ruimte in Regels, OCW 2017), designed to clarify what the law does and does not require – misconceptions clearly remain. While it is known for schools to 'overcompensate' when facing quality requirements with consequences attached (see Bevan & Hood 2006), this response seemed to be compounded by the unclear nature of these requirements in the Dutch context. The majority of participating schools followed prescribed methods, developed by companies in line with learning standards and test content, and the majority of teachers, as required by their management, were keeping careful yet time-consuming records of all aspects of their students' work should they be requested by the Inspectorate. This was reported to be a particular strain in schools that had recently undergone inspection. These practices transform the nature of school and teacher autonomy and more needs to be done to communicate to schools and boards, in a clear and transparent way, how their quality is (and how it is not) being assessed.

(Further) broaden quality measures: Over recent years, the Inspectorate's school quality framework has broadened. As well as focusing on quality assurance and learning outcomes, it also incorporates factors such as 'school climate' and 'educational process' (see Inspectie van het Onderwijs 2020b for latest framework). Yet risk-based inspection, determined by a school's average scores in the end-test, is still proving to be a problematic quality indicator. Given that school performance is closely linked to school population (see Diamond & Spillane 2004), and given that small group sizes can skew averages, it places unrealistic requirements and undue stress on particular schools. The 'low-performing' schools in this study, attended predominately by non-native Dutch students, were acutely aware of this quality indicator and felt under constant pressure to ensure sufficient results. Whilst the average score a school must achieve is adjusted according to its students' parental education level, participants in this study felt this adjustment to be insufficient. This puts an unfair burden on such schools, their teachers, and their students, and seems to result in a substantial focus on the core learning areas. Teachers described the ability to make regular time for non-core subjects (art, music, gym) as a luxury reserved for more privileged schools, and several considered their work pressures to be greater as a result of their school's population. This is substantiated by data showing 'the incidence of occupational burnout syndrome is higher among teachers in schools with a higher percentage of students from a non-Western migrant background…' (Inspectorate of Education 2019 p.33).

Since beginning school interviews, the Inspectorate has revised this indicator (Inspectie van het Onderwijs 2020a). From 2020/21, schools will not be measured on their average test score, but on the percentage of students attaining the basic level, and the percentage attaining the higher (target) level in the core learning areas. Further, this model will not only be adjusted according to parental education level, but also according to factors such as country of origin and length of stay in the Netherlands. These changes are an important step, recognising the unequal impact TBA is having across schools. It is hoped

that these new measures will reduce the number of schools, such as the one in this study, receiving inspections on the basis of (relative) underperformance while broader measures reflect a positive picture. Yet concerns remain. Firstly, the new indicators still exclusively focus on the core learning areas, meaning that underprivileged schools may still feel unable to take time away from numeracy and language to broaden their students' education. Secondly, although these are only minimum requirements, they will inevitably become the new yardstick by which school quality is measured, particularly as the aim is also for these indicators to be published. As a result, low and high performing schools alike may remain under considerable performance pressure. Risk-based indicators should be reconsidered, and first line quality measures should become broader, encouraging stakeholder perceptions of school quality, and schools' priorities to change.

Focus on the professionalism rather than the professionalization of teachers: For several years, and largely in an attempt to grow the profession, teaching in the Netherlands has been undergoing somewhat of a rebrand (see for example, Milner, Browes & Murphy 2020). The form that these efforts have taken reflect NPM values, including; encouraging learning from the private sector, increasing teacher certification and assessment requirements, and offering greater promotion opportunities (OCW 2013). Many teachers in this study saw these interventions as tokenistic, not offering the learning and development opportunities they sought. Rather than forcing change from the outside, while (accountability) structures that govern the profession remain unchanged, teachers should be given more space to change the profession from within. This would require a reduction in teachers' work tasks, thus making more time available for them to collaborate and learn from colleagues, shape and implement curricula and pedagogy to better meet students' needs, and follow professional development opportunities that are meaningful to them. This necessitates reduced surveillance, fewer performative and administrative demands, and a greater degree of trust to be afforded to schools and teachers, not only from the government but, significantly, from school boards and principals also. Voisin and Dumay's (2020) categorisation of different professional models, helps to provide ideas about alternative ways of structuring the Dutch teaching profession. This includes a move away from a 'market model' (shaped by standards-based regulation, relatively low levels of teacher autonomy and performance, market, and managerial forms of accountability), towards a more 'professional based' model (defined by characteristics such as the development of professional knowledge and skills, teacher autonomy and professional accountability). Indeed, given the high prevalence of teacher burnout and attrition, and the considerable teacher shortage in parts of the Netherlands (Inspectorate of Education 2019), an education model whereby trust in the professional is greater, and work demands fewer, seems to be needed sooner rather than later.

FINAL THOUGHTS

This thesis has aimed to shed light on the development and enactment of a global policy in a national context. Focusing on the Dutch case, it has further substantiated that strong similarities exist between the models of education governance being adopted in countries with vastly different institutional traditions, yet has also underlined the utmost importance of context in terms of how policy is shaped, how it is implemented, and what impacts it has. These contextual factors result in important differences to the formulations of TBA from one education system to another, yet nonetheless, the research has also shown that different formulations can still result in strikingly similar impacts. Although the SAWA model of education governance has been well-established for over two decades, its growth does not appear to be slowing down. It continues to be promoted by influential organisations such as the OECD, not only in its member states, but increasingly in non-member states and developing countries also. This is concerning, given the undesirable impacts of TBA that this study, and many others, have revealed. While such impacts are increasingly well documented, and formulations of policy tools are seemingly being adjusted in response (in the Netherlands, particularly since a new government came to power in 2017, plans to expand compulsory testing have been rescinded, and some TBA tools softened), the governance model itself, and the neoliberal principles on which it is based, show no signs of changing. The long-term impacts that this will have on education and teaching professionals remain to be seen. Can, for example, new teachers comfortably and sustainably adapt to this reality and the health of the profession improve? Or will they become increasingly weighed down by external demands that continue to sully the profession's attractiveness, thereby exacerbating the shortage of the workforce and teachers' work pressures? Beyond this, we must also consider the type of education being provided to students. The future shape of the labour market and of society in general is unknown, yet it seems clear at least that young people will require the ability and creativity to adapt, continuously learn, and apply their skills to a variety of problems. Education systems that incentivise a narrow focus through narrow notions of success, will struggle to equip students with such skills.

REFERENCES

Ball, S. J. (2016). Neoliberal education? Confronting the slouching beast. *Policy Futures in Education*, *14*(8).

Braun, A., & Maguire, M. (2018). Critical Studies in Education Doing without believing – enacting policy in the English primary school. *Critical Studies in Education*, *00*(00), 1–15.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. Public Administration, 84(3), 517–538.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality?. *Teachers college record*, *106*(6), 1145-1176.

Hammersley, M., & Gomm, R. (2008). Assessing the radical critiques of interviews. In M. Hammersley (Ed.), Questioning qualitative inquiry: Critical essays (pp. 89–100)). London: Sage.

Hendrikx, W., & Van Gestel, N. (2017). The emergence of hybrid professional roles: GPs and secondary school teachers in a context of public sector reform. *Public Management Review*, *19*(8), 1105-1123.

Inspectie van het Onderwijs. (2020a). Onderwijsresultatenmodel PO. Available at: https://www.onderwijsinspectie.nl/onderwerpen/onderwijsresultaten-primair-onderwijs/naar-een-nieuw-onderwijsresultatenmodel

Inspectie van het Onderwijs. (2020b). Onderzoekskader 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs.

Inspectorate of Education. (2019). *State of Education, 2020, Summary Report.* Utrecht: Inspectie van het Onderwijs.

Milner, A. L., Browes, N., & Murphy, T. R. (2020). All in this together? The reconstitution of policy discourses on teacher collaboration as governance in post-crisis Europe. *European Educational Research Journal*, *19*(3), 225–246.

Ministerie van Onderwijs, Cultuur en Wetenschap en de Inspectie van het Onderwijs. (2017) *Ruimte in Regels:Papieren rompslomp of kan het anders?* Den Haag.

Ministerie van Onderwijs, Cultuur en Wetenschap. (2013). Lerarenagenda 2013-2020: de leraar maakt het verschil.https://www.delerarenagenda.nl/documenten/publicaties/2015/01/01/lerarenagenda-oktober-2013

Noordegraaf, M. (2016). Reconfiguring professional work: Changing forms of professionalism in public services. *Administration & Society*, *48*(7), 783-810

Savage, G. C., & Lewis, S. (2018). The phantom national? Assembling national teaching standards in Australia's federal system. *Journal of Education Policy*, *33*(1), 118-142.

Voisin, A., & Dumay, X. (2020). How do educational systems regulate the teaching profession and teachers' work? A typological approach to institutional foundations and models of regulation. *Teaching and Teacher Education*, *96*, 103144.

Waslander, S., Hooge, E., & Drewes, T. (2016). Steering dynamics in the Dutch education system. *European Journal of Education*, *51*(4), 478-494.

Wilkins, C. (2011). Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system. *Professional Development in Education, 37:3, 389-409*

APPENDIX

TABLE 6: Data Collection & Participant Characteristics

RESEARCH QUESTION	DATA COLLECTED	PARTICIPANTS	RELEVANT CHARACTERISTICS
How has school accountability policy evolved in the Netherlands, and what role has school autonomy played within this?	Data collected between Oct 2017 - Feb 2018 1. 25 interviews with 31 policy experts 2. Analysis of policy documents linked to key pieces of (TBA) legislation: proposals of law, explanatory reports, parliamentary responses, advice from consultative bodies, letters from the Minister to the House of Representatives, final publications of law	Policy-maker 1	
		Policy-maker 2	
		Policy-maker 3	
		Policy-maker 4	
		Policy-maker 5	
		Academic 1	
		Academic 2	
		Academic 3	
		Academic 4	
		Academic 5	
		Policy Advisor 1	
		Policy Advisor 2	
		Policy consultant / teacher 1	
		Policy consultant / teacher 2	
		Policy consultant / principal	
		Policy worker, municipal level	
		Policy worker, municipal level	
		Rep. of Inspectorate of Education	
		Rep. of organisation for school boards (primary level)	
		Rep. of organisation for school boards (secondary level)	
		Rep. of national organisation for tests/exams	
		Rep. of testing agency	
		Rep. of national curriculum body 1	
		Curriculum policy designer 1	
		Curriculum policy designer 2	
		Rep. of teachers' union	
		Rep. of students' organisation	
		Rep. of parents' organisation	
		Rep. of teachers' movement / teacher	
		Rep. of large school board 1	
		Rep. of large school board 2	
What is the role of intergovernmental organisations in the promotion of SAWA and TBA?	Data collected between May 2019 – July 2019 1. Interviews with members of OECD policy review team and Dutch MoE 2. Content analysis of key documents: the 2016 RNPE, the Minister's invitation letter to OECD, ToRs of the review, the Minister's letter of response to the review	Policy review team, Dutch MoE	
		Policy review team, Dutch MoE	
		Policy review team, OECD	
		Policy review team, OECD	
		Policy review team, OECD	

In performative environments shaped by test-based accountability, how are teachers' practices and professional identities impacted?	Data collected between March 2019 - Nov 2019 1. 20 teacher interviews across 6 schools	Teacher 1, school 1	Group 8, female, < 5*
		Teacher 2, school 1	Group 7/8, male, < 5
		Teacher 3, school 1	Group 6, female, 10-14
		Teacher 4, school 1	Group 3, female, 10-14
		Teacher 1, school 2	Group 8, female, 10-14
		Teacher 2, school 2	Group 7/8, female, 15+
		Teacher 3, school 2	Group 6, female, 5-9
		Teacher 4, school 2	Group 5, male, < 5
		Teacher 1, school 3	Group 7, male, 10-14
		Teacher 2, school 3	Group 6, male, 10-14
		Teacher 3, school 3	Group 2, female, 10-14
		Teacher 1, school 4	Group 8, male, 15+
		Teacher 2, school 4	Group 8, female, 15+
		Teacher 3, school 4	Group 7, female, 15+
		Teacher 4, school 4	Group 5, male, 10-14
		Teacher 1, school 5	Group 8, female, 10-14
		Teacher 2, school 5	Group 7, female, 5-9
		Teacher 3, school 5	Group 6 (vice-principal at time of interview), female, 5-9
		Teacher 1, school 6	Group 8, female, 15+
		Teacher 2, school 6	Group 7, male, < 5
Is test-based accountability experienced differently by teachers in differently performing schools?	Data collected between March 2019 - Nov 2019 1. 20 interviews across 4 schools* (2 high, 2 low performing): 15 teachers, 4 vice/principals, 1 head of school board 2. Analysis of school documents: School guides	Director of School board	
		Vice-principal, school 1	High Performing school
		Teachers 1, 2, 3, 4, school 1	
		Principal, school 2	Low performing school
		Teachers 1, 2, 3, 4, school 2	
		Principal, school 2	Low performing school
		Teachers 1, 2, 3, school 3	
		Principal, school 4	High performing school
		Teachers 1, 2, 3, 4, school 4	
TOTAL		**61**	

*Number of years spent in teaching