The instrumentation of test-based accountability in the autonomous Dutch system

Browes, N.; Altinyelken, H.K.

Routledge
Taylor & Francis Group

Check for updates

The instrumentation of test-based accountability in the autonomous dutch system

Natalie Browes [a] and Hülya K Altinyelken [b]

aDepartament de Sociologia, Edifici B (Campus Bellaterra UAB), Universitat Autònoma de Barcelona, Barcelona, Spain; bDepartment of Child Development & Education, Universiteit van Amsterdam, Amsterdam, Netherlands

ABSTRACT

Test-based accountability or 'TBA,' as a core element of the pervasive Global Education Reform Movement (GERM), has become a central characteristic of education systems around the world. TBA often comes in conjunction with greater school autonomy, enabling governments to assess 'school quality' (i.e. test results) from a distance. Often, quality improvement is further encouraged through the publication of these results. Research has investigated this phenomenon and its effects, much of it focusing on Anglo-Saxon cases. This paper, drawing on expert interviews and key policy documents, couples a policy borrowing with a policy instruments approach to critically examine how and why TBA has developed in the highly autonomous Dutch system. It finds that TBA evolved incrementally, advancing towards higher stakes for schools and boards. Further, it argues that school autonomy has been central to the development of TBA in two ways. Firstly, following a period of decentralisation that increased school(board) autonomy, the Dutch government saw a need to strengthen accountability to ensure education quality. This was influenced by international discourse and accelerated by a (politically exploited) national 'quality crisis' in education. Secondly, the traditionally autonomous Dutch system, shaped by 'Freedom of Education', has at times conflicted with TBA, and has played a significant role in (re)shaping global policy and in mitigating the GERM.

1. 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, and Woessmann (2013) or Ko, Cheng, and 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 and 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 and Holloway 2019) is changing what education quality means, what is valued, and ultimately, changing teaching practices and professional identities (Day 2002; Evetts 2003, 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, and 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 2012). 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 (1), section (2) presents the theoretical framework used to analyse the development of the GERM and TBA, with a focus on the dynamics of policy borrowing and policy instrumentation linked to it. Section (3) presents the methodology and (4) the national context. Results are then presented (5) in three sub-sections addressing each of the research questions outlined above. Finally, in the discussion (6) and conclusion (7), 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.

2. 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, and Santori 2017; Verger, Parcerisa, and Fontdevila 2019). 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, Parcerisa, and Fontdevila (2019) have succinctly summarised these main GERM principles and the role of standardised testing in each case (Table 1).

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, Parcerisa, and

Table 1. The role of national assessments within the GERM.

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, and Fontdevila (2019), pp. 5.

Fontdevila 2019). 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 publically 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 and 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, and Fromewick 2009; Day and 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 and Janssens 2007 for an overview). Wider, more secondary impacts of these issues include increasing inequalities between schools and students and a narrowing and instrumentalisation of education, as well as teacher depro-fessionalisation and job dissatisfaction (Evers and Kneyber 2015; Verger and Parcerisa 2017). Yet despite these negative effects, the implementation of prescribed learning standards, standardised tests and accountability, continue to grow (Verger, Parcerisa, and Fontdevila 2019). 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 with 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 and Le Galès 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 2017), instrument choice is also guided by acceptance. Acceptance is based on perceptions of appropriateness or 'legitimacy' (Capano & Lippi 2017), 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 2017).

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 and 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 who will benefit from the instrument and who will be disadvantaged (Lascoumes and Le Gales 2007; Kassim and Le Galès 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, Novelli, and Altinyelken 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.

3. Methodology

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 stake-holders 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 second (or more accurately, first) roles as teachers, and one as a school principal. Their dual roles help 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 min. 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 about participant selection, the interview protocol and the 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 to deepen understanding of TBA policies in general.

4. 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:

4.1. Freedom of education

In the early 1900s, 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) (Glenn and de Groof 2012). 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.

4.2. Decentralisation

Decades later, during a period of decentralisation that began in the 1980s under a Conservative (confessional-neoliberal) government (Sleegers and 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 the 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.[1] 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 and Karsten 2015; Vedder 2006).

4.3. 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 professionalization. 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 an 'end-test,' made by the public-private 'CITO' organisation.[2] This test was first used in 1968 as a meritocratic, class-independent way of streaming students into secondary education. Since the 1980s, 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 and are 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 1970s was a messy and overcrowded curriculum (Kuiper and 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 to what degree schools actually experience curricula autonomy.

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.

5. Findings: the development & evolution of TBA in the Netherlands

5.1. 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 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, and 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 2000s, 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[3] 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 levels were incorporated into the primary end-test, and for the first time in its history, the test was made compulsory.[4] 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 have 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 and 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'.

5.2. 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 counter-balance 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 reveal a widespread concern over the quality of Dutch education in the years prior to the introduction of TBA policies. This concern grew throughout the late 1990s and early 2000s 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.[5] 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.

> (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 worryingly 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)

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 apparent 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 is 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)

5.3. *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 in doing so, advancing and transforming 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,[6] 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, and 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 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 a 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 end-test. A concern in parliament and in the field of education more broadly over the rise of a 'testing-culture,' coupled with fears of a 'checkout culture'[7] 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.[8] 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 have 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 liberalization in 2014. In 2018, these providers numbered six.[9] 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:

> 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 U.K. or the U.S., 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.

6. 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 (traditions 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, Parcerisa, and Fontdevila 2019).

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 1960s) and the LVS-tests (introduced as a student-learning aide in the 1980s) 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 2017). 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 may have 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, and 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, and 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.

7. 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 under-standing 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 policies 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 auton-omy 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, dialec-tical relationship between education reform and teacher professionalism (see for example Helgøy and 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 com-plex, 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.

Notes

1. Students deemed to have certain social disadvantages or learning difficulties would receive extra funding.
2. 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.
3. Cards outlining the inspector's main findings.
4. This excluded immigrant students in the country for less than four years, and students expected to go on to (advanced) special education (Ehren and Swanborn 2012).

5. In particular, this alludes to a government reform known as the 'Studiehuis' which restructured student learning at the upper secondary level.
6. 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.
7. Students and teachers no longer applying themselves for the remaining school year following the test.
8. 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.
9. This dropped to five this year due to insufficient order numbers for one provider.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Natalie Browes is a PhD researcher at the Autonomous University of Barcelona. She is currently conducting research as part of the international 'ReformEd' project, where her work examines the adoption, development and enactment of school accountability policies in the Autonomous Dutch system. In particular, she focuses on the impact of test-based accountability policies on primary school teachers' beliefs and practices in different school settings.

Hülya K Altinyelken is a senior lecturer and researcher at the Child Development and Education Department of the University of Amsterdam. Her research interests cover a wide range of issues, including global education policies, curriculum change, child-centred pedagogy, reform implementation and teachers. She is currently involved in two research projects on Muslim youth in the Netherlands, looking at the pedagogy of mosque education as well as its influence on identity development and social integration.

ORCID

Natalie Browes http://orcid.org/0000-0002-4451-9192
Hülya K Altinyelken http://orcid.org/0000-0003-2178-0862

References

Altinyelken, H. K., and S. Karsten. 2015. "The Netherlands: Structure, Policies, Controversies." In *Education in the European Union: Pre-2003 Member States* (Education around the world), edited by T. Corner, 305–324. London: Bloomsbury Academic.

Ball, S. J. 2003. "The Teacher's Soul and the Terrors of Performativity." *Journal of Education Policy* 18 (2): 215–228. doi:10.1080/0268093022000043065.

Ball, S. J., C. Junemann, and D. Santori. 2017. *Edu.net. Globalisation and Education Policy Mobility*. London: Routledge.

Berryhill, J., J. A. Linney, and J. Fromewick. 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. doi:10.1111/j.1468-0491.2007.00343.x.

Bovens, M. 2007. "Analyzing and Assessing Accountability: A Conceptual Framework." *European Law Journal* 13 (4): 447–468. doi:10.1111/j.1468-0386.2007.00378.x.

Capano, G., and A. Lippi. 2017. "How Policy Instruments are Chosen: Patterns Of Decision makers' Choices." *Policy Sciences* 50 (2): 269-293. doi:10.1007/s11077-016-9267-8.

CNV Onderwijs & EenVandaag. 2018. "*Onderzoek: 'leraren Onder Druk Om Schooladvies'.*" 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. doi:10.1177/0013124501332007.

Day, C. 2002. "School Reform and Transitions in Teacher Professionalism and Identity." *International Journal of Educational Research* 37 (8): 677–692. doi:10.1016/S0883-0355(03)00065-X.

Day, C., and L. Smethem. 2009. "The Effects of Reform: Have Teachers Really Lost Their Sense of Professionalism?" *Journal of Educational Change* 10 (2–3): 141–157. doi:10.1007/s10833-009-9110-5.

"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).

de Wolf, I., J. Verkroost, and H. Franssen. 2017. "Supervision and Accountability in the Dutch System." In *The Dutch Way in Education*, 111–121. Onderwis Maak je Samen & De Brink Foundation.

de Wolf, I. F., and F. J. G. Janssens. 2007. "Effects and Side Effects of Inspections and Accountability in Education: An Overview of Empirical Studies." *Oxford Review of Education* 33 (3): 379–396. doi:10.1080/03054980701366207.

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., and M. Swanborn. 2012. Strategic data use of schools in accountability systems. *School effectiveness and school improvement*, 23(2), 257-280.

Emons, W., C. Glas, and P. Berding-Oldersma. 2016. "Rapportage Vergelijkbaarheid Eindtoetsen." file:///C:/Users/User/Downloads/rapportage-vergelijkbaarheid-eindtoetsen%20(1).pdf

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

Evers, J., and R. Kneyber. 2015. *Flip the System*. Published in Association with Education International, London: Routledge.

Evetts, J. 2003. "The Sociological Analysis of Professionalism. Occupational Change in the Modern World." *International Sociology* 18 (2): 395–415. doi:10.1177/0268580903018002005.

Evetts, J. 2009. "New Professionalism and New Public Management: Changes, Continuities and Consequences." *Comparative Sociology* 8 (2): 247–266. doi:10.1163/156913309X421655.

Fontdevila, C. (forthcoming). "Researching the Adoption of School Autonomy with Accountability Reforms. A Methodological Note on Country Case Studies." *Reformed Methodological Papers. Number 1*. http://reformedproject.eu/tools/

Glenn, C., De Groof, J. 2012. *Balancing freedom, autonomy and accountability in education*. Wolf Legal Publishers (WLP).

Hanushek, E. A., S. Link, and L. Woessmann. 2013. "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA." *Journal of Development Economics* 104: 212–232. doi:10.1016/j.jdeveco.2012.08.002.

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. doi:10.1080/713698714.

Helgøy, I., and A. Homme. 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. doi:10.2304/eerj.2007.6.3.232.

"Inspectorate of Education, Ministry of Education, Culture & Science." 2018. *Summary report: The state of education 2016/17* (Eng.). file:///C:/Users/User/Downloads/116806_IvhO_The+State +of+Education_Hfst1_TG.pdf

Karsten, S. 1999. "Neoliberal Education Reform in the Netherlands." *Comparative Education* 35 (3): 303–317. doi:10.1080/03050069927847.

Kassim, H., and P. Le Galès. 2010. "Exploring Governance in A Multi-level Polity: A Policy Instruments Approach." *West European Politics* 33 (1): 1–2. doi:10.1080/01402380903354031.

Kauko, J., and M. Salokangas. 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. doi:10.1002/berj.3184.

Ko, J., Y. C. Cheng, and T. T. H. Lee. 2016. The development of school autonomy and account-ability in Hong Kong: Multiple changes in governance, work, curriculum, and learning. *International Journal of Educational Management*, 30(7), 1207-1230.

Kuiper, W., and Berkvens, J. (Eds.). 2013. *Balancing Curriculum Regulation and Freedom across Europe*. CIDREE.

Lascoumes, P., and P. Le Gales. 2007. "Understanding Public Policy through Its Instruments: From the Nature of Instruments to the Sociology of Public Policy." *Governance* 20 (1): 1–21. doi:10.1111/j.1468-0491.2007.00342.x.

Le Galès, P. 2010. Policy instruments and governance. In M. Bevir *The SAGE handbook of governance* (pp. 142-159). London: SAGE Publications

Lewis, S., and J. Holloway. 2019 March. Datafying the teaching 'profession': remaking the profes-sional teacher in the image of data. *Cambridge Journal of Education*, 49(1), 35–51.

OECD. 2013. Chapter 4: School Governance, Assessments and Accountability in *What Makes Schools Successful? Resources, Policies and Practices (Volume IV)*. pp.127–164, PISA, OECD Publishing

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*." Paris: PISA, OECD Publishing.

OECD. 2016b. "Reviews of National Policies for Education: The Netherlands, Foundations for the Future."

Peters, B. 2002. "The Politics of Tool Choice." In *The Tools of Government. A Guide to the New Governance*, edited by L. M. Salamon, 364–402. New York: Oxford University Press.

Rizvi, F., and B. Lingard. 2010. *Globalising Education Policy*. New York: Routledge.

Sahlberg, P. 2016. "The Global Educational Reform Movement and Its Impact on Schooling." In *The Handbook of Global Education Policy*, edited by K. Mundy, A. Green, B. Lingard, and A. Verger, 128–144. West Sussex: Wiley-Blackwell.

Scheerens, J. 2016. "The Netherlands: The Intersection of International Achievement Testing and Educational Policy Development." In *The Intersection of International Achievement Testing and Educational Policy: Global Perspectives on Large-scale Reform*, edited by L. Volante, 108–121. New York, NY: Routledge.

Schriewer, J. 1990. "The Method of Comparison and the Need for Externalization: Methodological Criteria and Sociological Concepts." In *Theories and Methods in Comparative Education*, edited by J. Schriewer and B. Holmes, 3–52. Bern: Lang.

Sleegers, P., and A. Wesselingh. 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*, 12(1), 20-33.

Steiner-Khamsi, G. 2014. "Cross-national Policy Borrowing: Understanding Reception and Translation." *Asia Pacific Journal of Education* 34 (2): 153–167. doi:10.1080/02188791.2013.875649.

Steiner-Khamsi, G. 2016. "New Directions in Policy Borrowing Research." *Asia Pacific Education Review* 17: 381–390. doi:10.1007/s12564-016-9442-9.

Teddlie, C., and F. Yu. 2007. "Mixed Methods Sampling: A Typology With Examples." *Journal of Mixed Methods Research* 1 (1): 77–100. doi:10.1177/1558689806292430.

UNESCO, & Global Monitoring Report. 2017. Netherlands Country case study prepared for the 2017/8 Global Education Monitoring Report *Accountability in education: Meeting our commitments*

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. doi:10.1080/03057920802447867.

Vedder, P. 2006. "Black and White Schools in the Netherlands." *European Education* 38 (2): 36–49. doi:10.2753/EUE1056-4934380203.

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., and L. Parcerisa. 2017. "A Difficult Relationship: Accountability Policies and Teachers. International Evidence and Key Premises for Future Research." In *International Handbook of Teacher Quality and Policy*, edited by M. Akiba and G. LeTendre, 241–254. New York, NY: Routledge.

Verger, A., C. Fontdevila, and L. Parcerisa. 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., L. Parcerisa, and C. Fontdevila. 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.

Verger, A., M. Novelli, and H. K. Altinyelken. 2012. Global education policy and international development: An introductory framework. *Global education policy and international development: New agendas, issues and policies*. Bloomsbury Publishing

Visscher, A., and M. Ehren. 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. doi:10.1080/15582159.2010.526845.

Waslander, S., E. Hooge, and T. Drewes. 2016. "Steering Dynamics in the Dutch Education System." *European Journal of Education* 51: 478–494. doi:10.1111/ejed.2016.51.issue-4.

Woessman, L. 2005. "The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-REPEAT and PISA." *In Education Economics* 13 (2): 143–169. doi:10.1080/09645290500031165.

World Bank. 2012. *SABER school autonomy and accountability country report: Netherlands 2012 (English)*. Systems Approach for Better Education Results (SABER) country report; 2012. Washington DC: World Bank. http://documents.worldbank.org/curated/en/570001468054259689/SABER-school-autonomy-and-accountability-country-report-Netherlands-2012

Yin, R. 2009. *Case Study Research, Design and Method*. Vol. 4. London: Sage Publications.