Comparative Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cced20

Outcome-based accountability regimes in OECD countries: a global policy model?

Björn Högberg & Joakim Lindgren

To cite this article: Björn Högberg & Joakim Lindgren (2021) Outcome-based accountability regimes in OECD countries: a global policy model?, Comparative Education, 57:3, 301-321, DOI: 10.1080/03050068.2020.1849614

To link to this article: https://doi.org/10.1080/03050068.2020.1849614

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

View supplementary material

Published online: 01 Dec 2020.

Submit your article to this journal

Article views: 930

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cced20

COMPARATIVE EDUCATION
2021, VOL. 57, NO. 3, 301–321
https://doi.org/10.1080/03050068.2020.1849614

Routledge
Taylor & Francis Group

OPEN ACCESS
Check for updates

Outcome-based accountability regimes in OECD countries: a global policy model?

Björn Högberg [a] and Joakim Lindgren [b]

[a]Department of Social Work, Umeå University, Umeå, Sweden; [b]Department of Applied Educational Science, Umeå University, Umeå, Sweden

ABSTRACT
The global diffusion of outcome-based accountability in education is contested, with accounts of universal convergence being challenged by perspectives emphasising heterogeneity across different national or local contexts. This study uses data from PISA to explore, firstly, the spatial and temporal diffusion of accountability across OECD countries, and secondly, whether accountability is implemented as a single coherent regime. Using cluster analysis techniques, we find that most countries fall into a 'Thick' accountability regime, with widespread use of most forms of accountability tools. However, this regime is not fully coherent, with some countries relying more on horizontal, and others on vertical, forms of accountability. A sizeable minority of countries fall into a 'Thin' regime, in which most accountability tools are largely absent. We also find indications of convergence across countries over time. We conclude that while accountability in education is indeed widespread, and increasingly so, it is not a universally dominant regime.

经合组织国家中结果导向的问责制度：一种全球政策模式？
结果导向的教育问责制在全球范围内的扩散备受争议，其有关趋同化的论述被强调国家或地区间多样性的观点所挑战。本研究利用国际学生能力测试（PISA）的数据，首先从空间和时间维度探究问责制在经合组织国家教育系统中的扩散情况； 其次探讨问责制是否作为一项单独、统一的机制来实施。利用聚类分析方法，我们发现大多数国家属于“厚”问责机制，即广泛应用多种形式的问责手段。但该机制并未呈现统一形态，有些国家更多依赖横向形式的问责制，而其他则倾向于纵向形式的问责制。一小部分国家属于“薄”问责机制，即缺乏大部分的问责手段。我们也发现，随着时间的推移，各国表现出趋同的迹象。我们的结论是，尽管教育中的问责制广泛存在，且日益普及，但并非一种普遍主导的机制。我们的结论是，尽管教育中的问责制广泛存在，且日益普及，但并非是一种普遍主导的机制。

KEYWORDS
Accountability; outcome-based accountability; PISA; education policy; policy diffusion; comparative education; cluster analysis; regimes

关键词
问责; 结果导向的问责; 国际学生能力测试; 教育政策; 政策扩散; 比较教育; 聚类分析; 机制

CONTACT Björn Högberg bjorn.hogberg@umu.se Department of Social Work, Umeå University, SE-901 87, Umeå, Sweden

Supplemental data for this article can be accessed https://doi.org/10.1080/03050068.2020.1849614.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Supranational organisations such as the Organisation for Economic Co-operation and Development (OECD) have been essential in bringing about a 'globalisation of education policy spaces', with policy tools, discourses and forms of governance being disseminated worldwide (Verger, Parcerisa, and Fontdevila 2019, 6; also Grek 2009; Lingard and Lewis 2016; Ramirez 2012). Through large-scale assessments, most notably the Programme for International Student Assessment (PISA), the OECD has contributed to a shift in education policy discourse, from a focus on content, resources and processes, to governance through outcome-based *accountability*, characterised by standardised measurement of performance, and the evaluation and incentivisation of this performance. Tellingly, the recent OECD publication *Governing Education in a Complex World* uses the term 'accountability' in various forms more than 500 times (Burns and Köster 2016).

However, key questions concerning the extent and characteristics of the global diffusion of outcome-based accountability remain contested and, as stated by a recent study on the topic, 'clearly, additional empirical work is needed to establish the global reach of accountability in education' (Sobe 2015, 144). In this study, we ask whether survey data from PISA can be used to shed new light on, firstly, the spatial and temporal diffusion of outcome-based accountability across OECD countries, and secondly, the extent to which alignment among different accountability policy tools allows us to describe accountability as a single coherent regime of education system governance.

We build upon and complement the pioneering work of Verger, Parcerisa, and Fontdevila (2019), who, partially drawing on PISA data, documented temporal trends in national large-scale assessments across OECD countries since the 1990s. Consistent with the main analytical technique used – cluster analysis – the study can be seen as largely exploratory. The ambition is to contribute to an empirically grounded discussion on the global diffusion of outcome-based accountability in education, in order to enable nuanced understandings of its diversity of national manifestations.

Theoretical background

Outcome-based accountability as a mode of governance

Several authors have pointed out that accountability is a slippery concept with different and partially overlapping meanings and connotations (Biesta 2004; Charlton 2002; Mulgan 2000; You 2017). The core meaning of the term refers to the process of being held accountable for something. Being held accountable implies that a person is obliged to provide an account of, for instance, their actions which, in turn, implies that there is an external party to whom this account is being provided. In this sense, accountability is a principal-agent type of social relation, between one party (the agent) who performs a certain activity, such as teaching, and a second party (the principal), who is a stakeholder in this activity. The social relation inherent in systems of accountability implies that the principal is not only a stakeholder but also has some form of authority or leverage over the agent, since it is the first party who is held accountable to the latter (Mulgan 2000). The account to which the agent is being held can refer to specific actions or inputs, or to the outputs or outcomes of these actions. Outcome-based accountability, evidently, is directed at the latter.

In contemporary discourse on education policy, the first party (the agent) is typically identified as the school, or as school personnel, while the second party (the principal) includes governmental bodies, but also parents and the general public (Ranson 2008). In an outcome-based accountability system, the account to which the school is being held is the academic performance of its pupils. It is this form of accountability that is the focus of this study.

Outcome-based accountability can be seen as a process in which three steps or dimensions can be distinguished. The first dimension concerns an emphasis on measurement of performance and the associated production of quantitative performance data (Biesta 2004). Since the second party (the principal) is external to the school, they need data that are standardised and comparable across pupils, teachers and schools (Charlton 2002; Troman 1989). The need to use performance data for evaluation implies a heavy use of standardised assessments and (high-stakes) testing, which, since the focus is on the end-product (the performance), are typically summative (Troman 1989).

The second dimension relates to how these data are used. Accountability, as stated, requires performance data to be evaluated by an external party. This external party is usually a governmental body or contracted organisation with delegated authority, but could also be parents or the general public (through, for example, public league tables) (Ranson 2008). The external evaluation is based on standards, criteria and benchmarks and is performed in accordance with standardised procedures in order to make schools commensurable and 'evaluate-able' (Charlton 2002; Holloway and Brass 2018; Ranson 2008). Performance data, school inspections or other forms of evaluation can also be strategically used by politicians, policymakers and other actors in processes of 'circulation', producing powerful feedback loops (Beer 2016).

This leads to the third dimension, which concerns the consequences, in the form of incentives, rewards or sanctions, of evaluations. The nature of these varies and can include market-based sanctions with 'parent-consumers' voting with their feet (Ehren and Perryman 2018) or administrative sanctions by, for instance, school inspectorates (Grek et al. 2013). The common denominator is that the evaluation has appreciable consequences for the agent being evaluated. Thus, the governing logic of accountability is the assumption that behavioural changes can be elicited from schools by attaching consequences to their measured performance (Ehren et al. 2015).

In sum, outcome-based accountability as a governance technique is constituted by the coupling of the following practices: (1) the production of standardised performance data with their (2) evaluation by external parties, and the (3) incentives and sanctions resulting from this evaluation. This definition may be considered as ideal-typical in a Weberian sense, i.e. as a theoretical accentuation of the core features of the phenomenon. An education system in which all three dimensions are salient and consistently aligned or coupled with each other would constitute an accentuated form of such a governance technique.

Comparative research on global policy models I: opposing perspectives

Outcome-based accountability, thus described, has its roots in public sector reforms in the Anglo-Saxon world, beginning in the era of Thatcher and Reagan in the 1980s, but has since become a paradigm for education reform worldwide (Biesta 2004; Ranson 2003). The literature on the global diffusion of accountability, and the role of international organisations and large-scale assessments in this process, is too broad and variegated to be covered in full here. Rather, in this study we ask which specific research problems can be fruitfully approached using cross-country survey data such as PISA. On the basis of this ambition, two contested questions appear to be particularly pertinent for further exploration:

1. The first question concerns the spatial and temporal diffusion of outcome-based accountability: How widespread is outcome-based accountability, especially in Western or industrialised countries, and are there signs of policy convergence over time?
2. The second question concerns the degree of alignment or coupling among different accountability policy tools: Are different tools implemented as parts of one coherent regime of education governance?

To structure the exposition of previous research on these questions, we draw on Schriewer's (2012, 416) distinction between, on the one hand, traditions or perspectives in comparative education emphasising an 'abstract universalism of trans-nationally disseminated models', and on the other, perspectives emphasising how 'this universalism fans out into multiform structural patterns wherever such models […] interact, in the course of their actual implementation, with different state-defined frameworks […]'. Such tensions between what may, in a somewhat simplified form, be labelled 'universalist' and 'contextualising' perspectives also shape the literature on the diffusion of accountability in education.

Regarding the first question, spatial and temporal diffusion, the arguably dominant account is of a global convergence of education systems within one 'world polity' (Kamens and McNeely 2010). Outcome-based accountability is understood to be a cornerstone of the 'Global Education Reform Movement' (Sahlberg 2016), a 'general narrative of transnational governance of […] education' (Erkkilä and Piironen 2014, 177), or of a 'Standard Global Reform Package' generating 'worldwide educational standardization' (Meyer and Benavot 2013, 9).

A leading theoretical basis for this 'universalist' perspective is neo-institutionalism and, more specifically, the world systems-framework (Ramirez 2012). Briefly, neo-institutionalist theory regards *legitimacy* and, by extension, experts and expertise, as essential for policy diffusion. Insofar as education systems share similar goals, such as learning, the presumption is that national education systems will embrace, at least at the level of discourse, similar 'best practices' or 'organisational blueprints' as they learn from the experiences of others. This is captured by the concept of *institutional isomorphism*: the 'growing cross-national similarity in […] education discourses, policies, structures and curricula' (Zapp and Ramirez 2019, 473). International organisations (e.g. OECD) and large-scale assessments (e.g. PISA) are in turn seen as key drivers of this diffusion process by conferring legitimacy on specific discourses and policies (Lewis 2017; Lingard, Martino, and Rezai-Rashti 2013; Meyer and Benavot 2013).

The 'universalist' account is challenged by scholars who acknowledge the importance of national or local contexts in mediating the impact of global diffusion processes and engendering varied forms of 'vernacular globalisation' (Lewis 2017; Maroy and Pons 2019; Müller and Hernández 2010; Ozga 2011). Such perspectives are often rooted in a broad theoretical canvas, including historical institutionalism, or in post-structural and postmodern traditions, and puts policy *implementation* at centre stage. While institutional isomorphism may hold at the level of discourse or paradigm (Ramirez 2012), the actual implementation processes may be more complex. National or local education systems are embedded in a web of interdependencies with other parts of society, constraining reform options and engendering path-dependent developmental trajectories. Likewise, conflicting interests among actors involved in policymaking may lead to selective or inconsistent adoption of 'best practices' (Schweisfurth and Elliott 2019). Thus, top-down accounts of the deployment of accountability are complemented by a focus on 'translation, hybridization, and bricolage' at different institutional levels (Maroy and Pons 2019, 13).

The tension between 'universalist' and 'contextualising' perspectives is carried over to the understanding of the second research question of this study: whether different policy tools are implemented as parts of one coherent model or regime. The focus of universalist, not least neo-institutionalist, accounts tend to be on how education systems strive to copy 'best practices' or organisational blueprints *in toto* from one context to another (i.e. isomorphism), leaving little room for, or de-emphasising, the importance of mediation at national or local levels (Schweisfurth and Elliott 2019). Thus, education systems adopt similar, and more or less coherent, *sets* or *combinations* of policy tools, with a high degree of alignment and tight coupling among different tools (cf. Erkkilä and Piironen 2014). In the context of this study, this would entail different countries adopting a similar, global accountability *regime*, in which the notion of regime refers to the alignment, or relations, among more concrete policy tools (cf. Ragin 1987).

Conversely, contextualising perspectives emphasise how universal models, such as the one previously described, are 'translated' in the process of implementation, and adapted to local conditions and contexts. Since these contexts vary along numerous dimensions, the universal model or paradigm of accountability can be translated into a range of idiosyncratic hybrid models (Maroy and Pons 2019). For instance, national case studies have revealed complex relations of accountability between actors, stakeholders and policymakers at various levels. A distinction is often made between vertical and horizontal relations of accountability, in which vertical relations are hierarchical and imply that the principal has formal power over the agent, while horizontal relations entail more decentralised decision-making, with the involvement of multiple stakeholders (Burns and Köster 2016). Which type that becomes dominant in a particular case may depend on the wider context in which relations of accountability are embedded.

In sum, we have two perspectives that paint partially conflicting pictures of the diffusion of outcome-based accountability: a universalist perspective stressing isomorphism and convergence towards one coherent regime; and a contextualising perspective stressing translation, mediation and hybridisation at various levels. Admittedly, this is a crude characterisation that hides nuances in these broad perspectives but it reflects the emphasis of the respective perspectives taken in their 'pure form'.

Comparative research on global policy models II: different methodological options

The universalist and contextualising perspectives discussed thus far are associated with different methodological traditions. The former is typically characterised by quantitative analysis of cross-national data (Ramirez 2012) or by analysis of policy documents and similar texts issued by international organisations (Erkkilä and Piironen 2014). The latter is characterised by detailed (often qualitative) national or local case studies, sometimes involving comparisons of two or three cases. The former tends to take a more generalising, variable-orientated stance, replacing the names of proper cases (e.g. countries) with variables representing specific properties of the cases abstracted from their context, or privileging universal policy paradigms over heterogeneity in the implementation of these paradigms. The latter takes a more particularising, case-oriented stance, paying closer attention to how the phenomena of interest are related to and embedded in distinct societal contexts (Schriewer 2006).

In this study we aim to explore a middle ground between these two. For this purpose, we use cluster analysis techniques and choose empirical indicators that intend to reflect the actual implementation of concrete policy tools. Cluster analysis is a generic term covering a family of methods that classify cases into groups (clusters) based on their degree of similarity with regard to a set of observed properties, with the groups constructed in such a way that cases within groups are as similar to each other as possible, while simultaneously being as dissimilar as possible to cases in other groups. Cluster analysis is an explorative technique, meaning that it builds the clusters from the bottom-up, as it were, based on the data, and is thus more inductive than deductive, and more data-driven than model-based.

Unlike the most common types of quantitative techniques, such as regression analysis, the focus of cluster analysis is not on relationships between variables, but on the cases themselves, or rather, on the relationships among different properties internally *within* cases (e.g. countries) and, simultaneously, on the relationships *between* cases. Cluster analysis can be thought of as a two-stage process whereby each case is first described as a distinct combination of certain properties, and these distinct combinations are then combined with other combinations, based on the similarities between their combinations of properties. Thus, a specific property (e.g. a given policy tool) is always embedded in web of interrelations with other properties. In this sense, cluster analysis is suitable for exploring ‘conjectured relationships’ or ‘relations of relationships’ (Schriewer 2006, 310), at least insofar as these relationships can be described quantitatively.

Cluster analysis retains the advantage of universalist and variable-oriented traditions of being able to analyse large quantitative datasets with many cases and properties (variables), while simultaneously avoiding more radical forms of abstract decontextualisation by analysing the properties as embedded in distinct interrelations. For this reason, cluster analysis has been extensively used in policy-orientated comparative research, for instance, in research on welfare state regimes (Castles and Obinger 2008). While to our knowledge it has never been employed in research on accountability, it is consistent with our focus on alignment among different policy tools in accountability ‘regimes’.

Data and methods

Data

We use data from OECD’s Programme for International Student Assessment (PISA). PISA is the largest and arguably most influential international school survey and has been instrumental in promoting outcome-based accountability worldwide (Grek 2009; Lingard, Martino, and Rezai-Rashti 2013). An advantage of PISA data is that the items intend to capture the actual adoption of specific policy tools. This is important since much previous comparative research on accountability has tended to focus on discourse and policy paradigms, thus running the risk of neglecting variation in how these paradigms are implemented (Maroy and Pons 2019, 88).

PISA contains a school background questionnaire with questions about the policies and practices of the school, distributed to the school principal. Data are collected through a two-stage sampling design, with schools as the primary sampling unit, and constructed in order to be representative of the 15-year-old pupil population in each country (OECD 2017). With some exceptions (see Table S2 in the supplemental online material) we use data from the 2015 PISA survey. We refrain from using the most recent PISA wave from 2018 due to the poorer availability of suitable indicators. The last stage of the analysis is complemented by previous PISA surveys in order to track the diffusion of accountability tools over time. We limit the analysis to OECD countries in order to make the very large amount of data more manageable, resulting in a set of 35 country cases.

Included variables

We distinguish between three dimensions of outcome-based accountability, which we aim to capture through a variety of indicators (variables). Most of the included indicators are binary, measured by statements or questions with only ‘Yes’ or ‘No’ as response options. Thus, the indicators show the average proportion of pupils in a country who attend schools in which the principal reports that the statement is true.

The first dimension concerns standardised measurement of performance and, in this study, is measured by (1) whether or not the school systematically records performance data (test results and graduation rates), and reports that this is mandatory according to governmental policies; and (2) the frequency of standardised testing at the school. The second item has five response options: ‘Never’, ‘1–2 times per year’, ‘3–5 times per year’, ‘Monthly’ or ‘More than monthly’. We transform this into two indicators, one measuring the proportion that respond ‘Never’, and one measuring the proportion that respond either ‘3–5 times per year’, ‘Monthly’ or ‘More than monthly’.

The second dimension concerns the use of external evaluation in relation to performance data. Evaluation by governmental bodies is measured using an item that asks whether ‘Achievement [performance] data are tracked over time by an administrative authority’; evaluation by parents by an item asking whether ‘Achievement [performance] data are provided directly to parents’; and evaluation by the general public using an item asking whether ‘Achievement [performance] data are posted publicly (e.g. in the media)’. Moreover, due to the centrality of comparison for external evaluation, we include one item that asks whether standardised tests are used to ‘compare the school to district or national performance’.

The third dimension concerns the consequences of evaluations. We measure this using one item that asks whether external evaluation aimed at quality assurance and improvement is mandatory. This does not measure sanctions directly but can be seen as a proxy for the influence of school inspectorates and similar bodies. We also include an item that asks whether there is pressure on the school from parents to ‘set very high academic standards and to have students achieve them’, with the following response options: ‘There is constant pressure'; 'Some pressure', 'Pressure is largely absent'. We transform this into two indicators, one measuring the proportion that respond 'There is constant pressure' and the other the proportion that respond that 'Pressure is largely absent'. These capture the extent to which schools are held accountable to parents, thereby coming close to measuring the extent of market-like sanctions.

The three dimensions are measured using three or four indicators each. This is important since cluster analysis gives each indicator equal weight. In addition to these indicators, we include one item measuring whether standardised tests are used to 'make judgements about teachers' effectiveness'. This can be seen as indicating the degree to which teachers are held accountable for the performance of their pupils.

In varying degrees, the indicators capture horizontal and vertical relations of accountability. Indicators that primarily capture horizontal relations, i.e. with parents or the general public, are the provision of performance data directly to parents, public posting of performance data, and pressure from parents. The use of standardised tests to compare the school to district or national performance can also be seen in this context, as it encourages lateral competition between schools. Indicators that primarily capture vertical relations include the provision of achievement data to administrative authorities, mandatory external evaluation, and the use of tests to make judgements about teachers' effectiveness (implying vertical relations of authority in schools).

All items, including descriptive data, are listed in Tables S1 and S2 in the supplemental online material.

Data analysis

In the first stage of the analysis, we generate indicators based on the country averages of all the previously described items. Specifically, we estimate the average proportion of each country for each possible response option for the respective items. Based on the averages, we then generate a new data set, with countries as units, and each country is given equal weight in the analysis.

As previously described, we use (hierarchical) cluster analysis to examine interrelations of accountability policies within and between countries. In hierarchical cluster analysis, all cases begin by forming their own 'cluster', and are then linked successively to each other based on their degree of similarity across all indicators (variables), until, ultimately, all cases belong to one and the same cluster. The term *hierarchical cluster analysis* implies that the clusters are hierarchically related, with all lower-level clusters included in the higher-level clusters. Due to the exploratory purpose of the analysis, we use agglomerative hierarchical cluster analysis, in which the number of clusters is not defined beforehand. As is common in comparative policy research (Castles and Obinger 2008), we use Ward's linkage method, with Euclidean distance, to form the clusters.

Results

Regimes of accountability

We begin with the classification of countries in Figure 1. In cluster analysis, there is no unique solution to determine when the clustering is 'optimal' and analyses of the

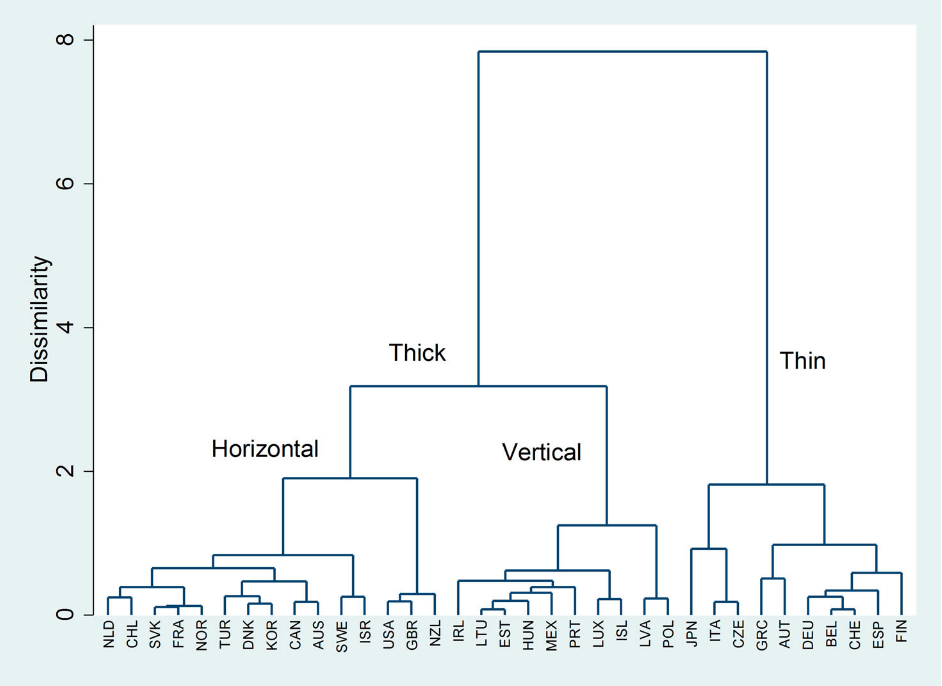

Figure 1. Dendrogram depicting the formation of clusters.

number and distinctiveness of specific clusters are best seen as acts of interpretation. Figure 1 depicts a so-called dendrogram, or cluster tree, which is a graphical illustration of the different steps of the clustering process. The bottom of the cluster tree depicts the starting point, in which all cases (countries) form their own cluster. Each vertical step then displays the stage at which a case or cluster is linked to another one until, at the top, all cases are grouped together in one cluster. The point at which cases are linked shows their relative distance: the earlier they are linked, and the shorter the vertical lines, the more similar they are.

The most distinctive clustering, in the sense of where the vertical lines are the longest, is the last one, in which two clusters are separated. At this stage, all countries, save for 10, are joined in one large cluster. The second and smaller main cluster (labelled 'Thin' accountability regime; see below for explanations of the labels) is internally rather homogenous, and comprises Japan, Italy, Czech Republic, Greece, Austria, Germany, Belgium, Switzerland, Spain and Finland. However, in the first and larger of the two main clusters (labelled 'Thick' accountability regime), two sub-clusters can be identified. One sub-cluster (with the provisional label 'Horizontal') comprises The Netherlands, Chile, Slovakia, France, Norway, Turkey, Denmark, Korea, Canada, Australia, Sweden, Israel, the USA, Great Britain and New Zealand. The other sub-cluster ('Vertical') comprises Ireland, Lithuania, Estonia, Hungary, Mexico, Portugal, Luxembourg, Iceland, Latvia and Poland. The Thin regime is dominated by continental (including southern) European countries, while the Thick regime is more heterogeneous, although it is noteworthy that all Anglo-Saxon countries, all Nordic countries save for Finland, and all eastern or central European countries, save for the Czech Republic, are included here. A notable feature is also the

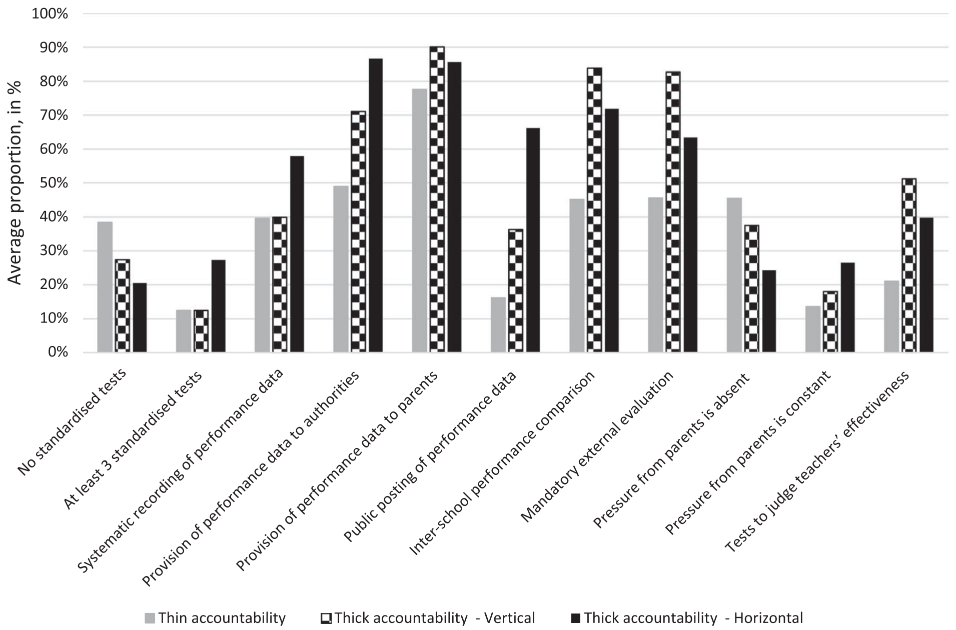

Figure 2 . Regime averages for indicators of accountability policy tools. Data from PISA 2009–2015. Note that *lower* values on 'No standardised tests' and 'Pressure from parents is absent' indicate *more intense* use of accountability tools.

closeness of the two archetypical accountability countries, the USA and Great Britain. Together with New Zealand (another 'early adopter' (Willis 1992)), they form a distinct subgroup until quite late in the clustering process.

So what are the characteristics of the clusters? We present the within-cluster averages for the respective indicators in Figure 2, and a summary of the characteristics of the regimes in Table 1. While a two-cluster solution fits the data best, we also present the averages for the two sub-clusters previously described (Horizontal and Vertical). Beginning with the characteristics of the Thin regime (grey bars), we see that it has a consistently and sometimes substantially lower use of accountability tools than the two sub-regimes in the Thick regime. For instance, on average, 16% of pupils in the Thin regime

Table 1. Summary of regime characteristics.

Regime	Countries	Production of performance data	Evaluation by external parties	Consequences of evaluations	Horizontal vs. vertical
Thick accountability regime – Horizontal	Anglo-Saxon countries Scandinavia Mostly high-income	High	High	High	Relatively stronger horizontal accountability
Thick accountability regime – Vertical	Central/Eastern Europe Mostly middle-income	Medium	High. Low for public posting of achievement data	Medium or high	Relatively stronger vertical accountability
Thin accountability regime	Continental Europe Mostly high-income	Low	Low. High for the provision of achievement data to parents	Low	Neither horizontal nor vertical accountability

attend schools in which performance data are publicly posted, compared to 36–66% in the Thick regime. There are also significant differences regarding the frequency of standardised testing, parental pressure and inter-school performance comparison. Between-regime differences are smaller regarding the provision of performance data to parents and the systematic collection of performance data, suggesting that countries in the Thin regime still collect such data, but rarely use them for incentivising schools.

The overall picture emerging from the differences between the two main clusters, the Thick and Thin regimes, is that accountability can be described as 'more' or 'less' along one continuum, with countries in the Thick regime characterised by a greater use of most types of accountability tools. This is indicative of some degree of alignment and coupling among different tools, as the use of one tool (e.g. standardised testing) tends to go hand in hand with the use of other tools (e.g. performance reporting).

When considering the differences *within* the Thick accountability regime, between what we label the Horizontal (black bars) and Vertical (chequered bars) sub-regimes, things are slightly more complex, and the difference, though less marked, is between different *forms* of accountability. The Horizontal sub-regime scores higher on most indicators of horizontal relations – public posting of performance data (66% vs. 36%) and strong parental pressure (26% vs. 18%) – as well as on the frequency of standardised testing. The Vertical sub-regimes score higher on indicators of vertical relations – mandatory external evaluation (83% vs. 63%) and the use of standardised tests to judge teachers' effectiveness (51% vs. 40%). One interpretation of this is that accountability in the Horizontal sub-regime is based on incentives, competition and decentralised forms of control, while modes of bureaucratic control or other forms of hierarchical evaluation play a relatively more prominent role in the Vertical sub-regime. Note, however, that vertical relations (e.g. mandatory external evaluation) are by no means non-existent in the Horizontal sub-regime. This may be seen as indicative of the 'odd combination of marketisation on the one hand and centralisation of control on the other', noted by Apple (2005, 11), i.e. that market reforms are often followed by attempts to regain centralised control through, for instance, audit and inspection.

Regarding the individual countries, the Horizontal sub-regime includes all Anglo-Saxon countries, save for Ireland, while the Vertical sub-regime includes more eastern or central European countries, and relatively more middle-income countries (exceptions are Ireland, Iceland and Luxembourg).

The differences between the Horizontal and Vertical sub-regimes suggest that the context in which accountability is embedded may influence the kinds of policy tools that become more salient. Of particular importance in this regard is the extent to which accountability is related to the degree of school marketisation and autonomy; two policy domains that are often highlighted as being parts of the same reform model and policy paradigm, but that may be more easily aligned with horizontal forms of accountability (Erkkilä and Piironen 2014; Sahlberg 2016).

We explore this issue by looking at three PISA indicators of private or market elements in education: the proportion of pupils in private schools, the proportion of private funding of the school, and the proportion of pupils in schools in which there is more than one other school that is competing for pupils in the area. We also look at two indicators of school autonomy: the proportion of pupils in schools in which the school has considerable responsibility for budget and employment decisions,

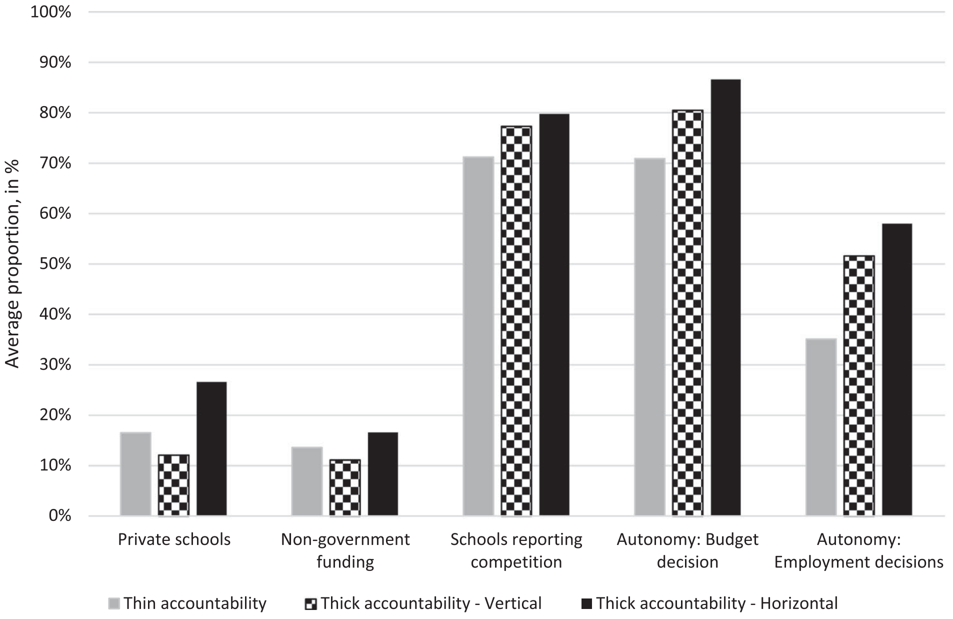

Figure 3 . Regime averages for indicators of market elements and school autonomy. Data from PISA 2012–2015.

respectively (see Table S1 in the supplemental online material for more details of the indicators). The results, presented in Figure 3, show that across all five indicators, the degree of market elements and autonomy is highest in the Horizontal sub-regime, and lowest in the Thin regime; the exception is the relatively high proportion of private schools in the Thin regime, which reflects the tradition of religiously-run schools in many continental European countries. The Vertical sub-regime generally falls in between, but closer to the Horizontal sub-regime. Overall, consistent with 'contextualising' perspectives, these results suggest that the way in which accountability is related to other policy domains is important for understanding how specific tools are implemented, as well as what effects these tools have (cf. Verger, Moschetti, and Fontdevila 2020).

Temporal trends in accountability

In the final stage of the analysis, we investigate trends in accountability from 2000 to 2015. As previously mentioned, universalist perspectives tend to describe a convergent process, whereby more and more countries 'catch up' with the best practices of early adopters. Contextualising perspectives, on the other hand, describe a process in which change is slow, heterogeneous and non-linear, in the sense that not all education systems move steadily in the same direction (Maroy and Pons 2019; Müller and Hernández 2010).

Figures 4 and 5 show trends across countries for the two indicators that have been included in almost (not 2006) all PISA waves since the first wave in 2000: inter-school performance comparisons and the use of tests for judging teachers' effectiveness. In this

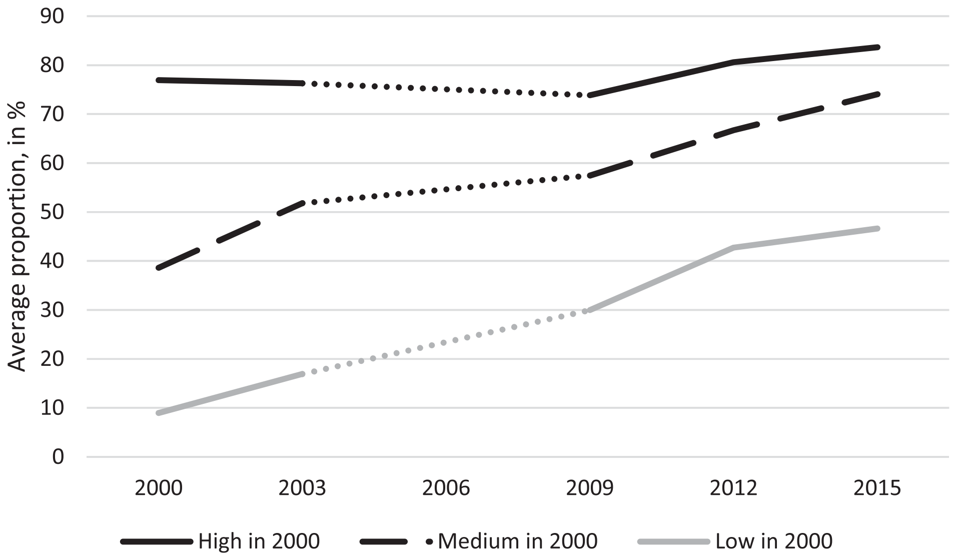

Figure 4. Proportion of pupils in schools that compare the school to district or national performance, 2000–2015. Data from PISA, waves 2000, 2003, 2009, 2012, 2015. Gaps are marked by dotted lines.

analysis, we cannot use the regimes generated by the cluster analysis since these are based on more recent data and their existence should not be extrapolated back in time. Instead, we ranked all countries based on their use of the respective policy tool in 2000 (that is, their score on the respective indicator in 2000) and grouped them into three equally sized groups: low, medium and high, with the 33rd and 67th percentiles as cut-offs. The figures depict the average proportion in each country group that uses

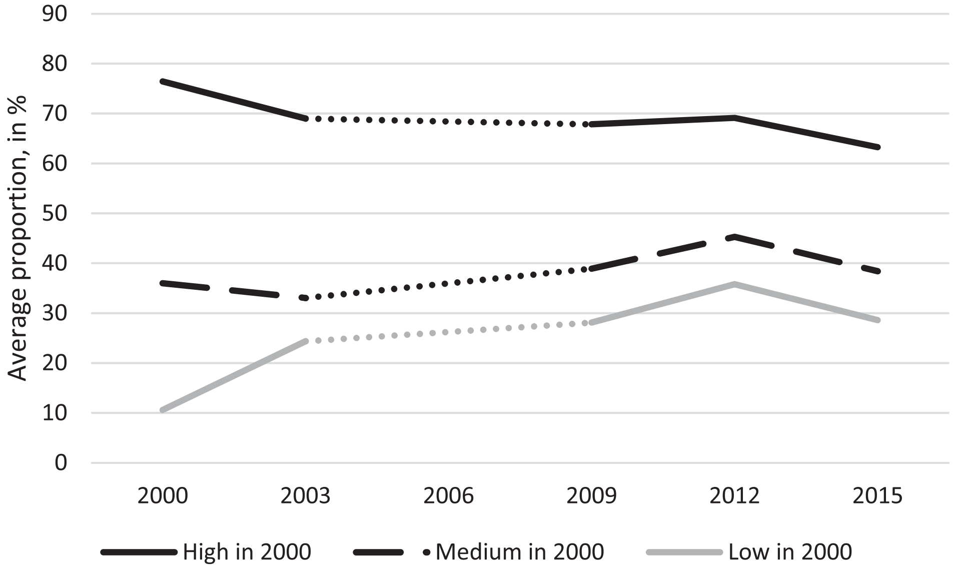

Figure 5. Proportion of pupils in schools that use tests to judge teachers' effectiveness, 2000–2015. Data from PISA, waves 2000, 2003, 2009, 2012, 2015. Gaps are marked by dotted lines.

the respective policy tool (vertical axis), and how this has changed from 2000 to 2015 (horizontal axis). Solid black lines show the levels in the country group with initially high use in 2000, dashed black lines the equivalent in the medium group, and solid grey lines the equivalent in the low group. Dotted lines indicate missing data in 2006. Since the groups are formed based on country scores in 2000, the composition of countries in the three groups is the same across all years, but they vary across the two indicators. Note that Chile, France, Israel, Slovakia, Turkey, Estonia and Lithuania did not participate in all survey waves and are therefore excluded here.

Inter-school performance comparisons (Figure 4) increased in all country groups, but more so in countries with initially low or medium use of this tool, leading to a converging pattern in 2015. The stable development in the group with initially high use is probably partially due to ceiling effects, since the proportion was already approaching 100% in 2000. As for the use of tests for judging teachers' effectiveness (Figure 5), this actually decreased in the group with initially high use, while increasing quite significantly in the low group, and increasing slightly in the medium group, yet again leading to convergence (though not a universally higher prevalence) in 2015.

While it is difficult to draw firm conclusions based on only two indicators, the impression from Figures 4 and 5 is that, in line with the findings of Verger, Parcerisa, and Fontdevila (2019), accountability has become more widespread, but more so in countries in which it was initially less common, leading to convergence across countries. There are also signs of convergence in the use of standardised tests, but due to larger gaps in the availability of data over time, we view these trends as being less reliable and do not show the results here.

Sensitivity analyses

Results of statistical models are always sensitive to choices made regarding what data to use and how to operationalise indicators. With binary indicators, the operationalisation is straightforward, but with indicators that include many response options, i.e. frequency of standardised tests, it was necessary to merge response options in order to generate more tractable indicators. To probe how sensitive the results are to this choice, we re-estimated the cluster models but replaced the two indicators 'No' and 'At least three' tests with a single indicator representing the weighted average of all five response options. The results were largely similar, with two main clusters (Thick and Thin regime) emerging, and the same countries placed in the respective main clusters. However, within the Thick regime, Korea and Chile both ended up in the Vertical instead of the Horizontal sub-regime, which testifies to the comparatively less robust nature of these sub-regimes.

In the main analysis we also chose to exclude the most recent data, from 2018, since only seven out of 11 indicators were included in this wave. We have re-estimated the cluster models with the latest (2018) data on these seven indicators. The available indicators for this analysis are systematic recording of performance data, the provision of performance data to administrative authorities and to parents, public posting of performance data, inter-school performance comparisons, mandatory external evaluation, and the use of tests to judge teachers' effectiveness. Again, the clustering of countries was largely similar, with two main clusters, the Thick and Thin regimes, emerging. However, Korea switched to the Thin regime, and Italy to the Thick regime (and Vertical sub-regime). Within the Thick regime, Mexico and Portugal switched to the Horizontal sub-regime.

One conclusion from these sensitivity analyses is that while the overall regime pattern appears to be relatively robust, in the sense that cross-country variation is substantial and not all education systems converge on one dominant model, the classification of specific countries is contingent on the choice of data and the operationalisation of indicators.

Discussion

The present study set out explore two questions concerning the diffusion of outcome-based accountability. (1) How widespread is outcome-based accountability, and are there signs of policy convergence over time? And (2) are different accountability policy tools implemented as parts of one coherent regime of education governance? Though we view the study as being primarily exploratory, some tentative conclusions can be drawn on the basis of this analysis.

With regard to the first question, our findings indicate that accountability tools are quite widely used in most OECD countries, with a majority of countries falling into what we labelled a Thick accountability regime. Moreover, though the sparseness of data prohibits firm conclusions being made regarding time trends, we found signs of convergence across countries in their use of accountability tools between 2000 and 2015. Nonetheless, diffusion and convergence are not all-encompassing, with substantial variations in the degree of policy implementation across countries, and a sizeable minority of countries falling into what we labelled a Thin accountability regime.

With regard to the second question, we found a fairly high degree of alignment and coupling among different accountability tools, in the sense that the use of one tool tends to go hand in hand with the use of others. Thus, countries could be grouped into two main clusters along one continuum, the Thick and Thin accountability regimes, with almost all policy tools being more frequently used in the former. In this sense, accountability may be described as one quite coherent model, or regime. However, we also found indications of two sub-regimes within the Thick regime, with one sub-regime (Horizontal) characterised by relatively more decentralised relations, and the other (Vertical) based on hierarchical forms of control. While these sub-regimes were less clear-cut and robust, they suggest that important but subtle differences exist concerning which specific policy tools dominate in specific cases.

The country classification conforms to some previous comparative studies. For instance, the salience of the USA, Great Britain and New Zealand in the Thick accountability regime is consistent with the notion that ideological transformations in these countries were instrumental in bringing accountability to the forefront of global education reform (Biesta 2004; Ranson 2003; Willis 1992). The dominance of continental European countries in the Thin regime is also consistent with previous findings (Müller and Hernández 2010; Verger, Parcerisa, and Fontdevila 2019). Moreover, the distinction between horizontal and vertical accountability resembles similar distinctions made by other scholars, for instance, between market and administrative accountability (Verger, Parcerisa, and Fontdevila 2019), market-based and government-based accountability (Harris and Herrington 2006), or consumer and performative accountability (Ranson 2003).

Before moving on to discuss the implications of these findings, we want to stress some important limitations of the study that readers should bear in mind when interpreting them. Classificatory approaches such as this may become static, and lead to a reification of a given, more or less historically contingent, snapshot of the world. Strictly speaking, the empirical country clusters, or regimes, are only groups of cases that share similarities according to the measured properties, although there is always a risk of conceptual stretching when theoretical concepts are used to describe actual cases. Thus, it should be kept in mind that the identification of the clusters as 'regimes' is fundamentally an act of interpretation and labelling. Furthermore, we could not account for variations within national education systems, nor investigate the vast majority of countries in the world that are not members of the OECD, despite the continuing expansion of PISA and OECD:s 'system of governance' into low- and middle income countries (Xiaomin and Auld 2020).

Implications for comparative education

The comparative literature on accountability in education can be sorted into what we – drawing on Schriewer (2012) – labelled 'universalist' and 'contextualising' perspectives, with an arguably dominant universalist perspective that stresses diffusion, convergence and isomorphism being challenged by contextualising scholars emphasising varied forms of 'vernacular globalization', mediation, embeddedness and hybridisation.

What are then the implications of this study with regard to these theoretical perspectives or traditions? On the one hand, the results point towards a relatively high degree of international diffusion and even isomorphism, with a majority of OECD countries adopting a moderately coherent regime of 'Thick 'accountability, and with signs of convergence across countries over time. Moreover, the fact that diffusion and convergence are visible from the PISA data themselves, and after the launch of PISA in 2000, suggests that international organisations and testing programmes may play a part in the policy diffusion process, in line with previous accounts (Grek 2009; Lingard, Martino, and Rezai-Rashti 2013; Ramirez 2012). On the other hand, the diffusion of the 'Thick' accountability regime is by no means universal, as testified by the sizeable minority of countries adopting a 'Thin' regime of accountability. The distinct characteristics of these countries suggest that the predominant focus on Anglo-Saxon education systems in the literature on accountability may disregard important heterogeneity in adoption and implementation across different contexts (Maroy and Pons 2019). Moreover, the existence of sub-clusters or sub-regimes indicates that any ostensibly universal model of accountability is not fully coherent, and the relations between accountability regimes and market elements and school autonomy demonstrate the importance of viewing accountability as embedded in larger educational contexts. The persistent diversity of national education systems, and their clustering into more or less distinct regimes, is consistent with recent research by Johansson and Strietholt (2019). Drawing on arguments from historical institutionalism, they find that countries' achieved curricula in mathematics are not converging, that regions with shared traditions (e.g. Post-Soviet states) continue to cluster together, and that differences in curricula are deeply rooted in historical and cultural contexts. All in all, we find support for both 'universalist' and 'contextualising' perspectives, and on the basis of these findings we cannot conclude that either provides a more valid account of the diffusion of accountability in education.

The findings of the study may moreover cast some new light on the relationship between accountability and learning or educational performance. At least at the level of public discourse, the arguably dominant motivation behind accountability reforms has been to improve measures of educational performance, and even to improve countries' rankings in international league tables (e.g. PISA) (Auld and Morris 2014; Lundahl, Hultén, and Tveit 2017). Against this background, it is striking that there is no clear-cut association between the use of accountability tools and tests-results in PISA. The Thin regime, in which accountability plays a minor role, includes several countries typically regarded as high-performing, such as Finland, Japan, Switzerland and Germany (and Korea, when using data from 2018), while the Thick regime, with more widespread use of accountability tools, includes several relatively low-performing countries, such as Luxembourg, Iceland, Chile and Israel (OECD 2016). Moreover, there is no straightforward association between accountability and social equality in educational performance, as measured in PISA (OECD 2016). Finland and Japan (and Korea) combine high equality with thin accountability, while some thin-accountability, Germanic countries (e.g. Austria and Germany) are quite unequal, probably due to their highly tracked education systems. Likewise, some Nordic countries (e.g. Denmark and Norway) combine high equality with thick accountability, while more neoliberal thick-accountability countries tend to be more unequal (e.g. New Zeeland and Chile).

While further research is clearly required to disentangle these complex associations, a tentative conclusion may be that accountability tools are comparatively inconsequential for the kinds of measures that are in focus in international large-scale assessments, despite the fact that these measures are often put forth as motivations for accountability reforms in the first place. We argue that these results demonstrate the critical potential of comparative education – how innovative analytical approaches to large-scale assessment data can raise questions and challenge accepted ideas and widely held policy assumptions.

As for the methodological implications, we opened this study by asking whether survey data from PISA can be used to shed new light on the question of global outcome-based accountability. We also asked whether cluster analysis, viewed as a middle ground between, on the one hand, universalist and variable-orientated, and on the other, contextualising and case-orientated, traditions in comparative research, can be productively applied in analysing these data. Others may judge whether this approach has indeed brought useful insights in this particular case. Here, we would like to end by discussing how the approach, and the empirical results produced by it, may inspire future comparative research.

Concerning the overall approach, or research design, we see opportunities for using cross-national survey data, including large-scale assessments, to explore other issues of relevance to comparative education. These may include the experiences, identity and well-being of teachers or pupils, and how these interact with education policies. Relevant data, with teachers or students being surveyed, can be found in PISA, as well as in the Teaching and Learning International Survey (TALIS) or the Health Behaviour in School-aged Children survey (HBSC). We also believe that educational researchers attuned to the embeddedness of education policy in local contexts may benefit from exploring quantitative approaches that are less abstract and de-contextualising, such as data on policy implementation, or cluster analysis. Johansson and Strietholt (2019), combining case- and variable-oriented approaches through a latent class analysis, is a recent example of how PISA data can be used for comparative purposes without losing sight of contextual particularities. The arguments put forth by Elliott et al. (2019), focused on how cross-national survey data can be used for contextualised analysis within an 'ecosystemic' framework, are also relevant in this regard.

Concerning the empirical results, we again stress the explorative nature of the analysis and caution against reifying the specific classification of the countries that we present. Nonetheless, we believe that future comparative work may find the results relevant in several ways. Firstly, the regimes, or more generally the data and indicators used to construct them, may serve as *explananda* in their own right, and help guide further research on how, why and under what conditions different education systems adopt, or do not adopt, different regimes of accountability. In this regard, the notion of regimes can be aligned with analyses of the role of different political contexts for policy diffusion (cf. Esping-Andersen 1990), while cross-national time series data (e.g. PISA) can be combined with a focus on the importance of international organisations. Secondly, the regimes may also serve as *explanans*, as theoretical and empirical tools to understand the consequences of different regimes of accountability in other educational domains, including the consequences for teachers and pupils, or the consequences for learning outcomes. Thirdly, to the extent that a classification of countries like this is deemed empirically and theoretically relevant, the resultant classification may help comparative researchers relate case studies from specific national settings to each other. Cluster analysis as a method, and regimes as a concept, take aim at the relationships between phenomena (e.g. policy tools), or even at the 'relations of relationships' (Schriewer 2006), thereby bringing some 'translatable' institutional context to the included cases. In this sense, relevant country classifications may expand the universe of relevant cases for non-quantitative comparative research.

Acknowledgements

The work on this article (planning, analysis and writing) was funded by VR; the Swedish Research Council (grant number 2018-03870_3).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work on this article (planning, analysis and writing) was funded by VR; the Swedish Research Council (grant number 2018-03870_3).

Notes on contributors

Björn Högberg is a researcher at the Department of Social Work, Umeå University. The present study is part of a research project investigating the consequences of education reforms for the well-being of pupils. Orcid: 0000-0002-0199-0435

Joakim Lindgren is an Associate Professor at the Department of Applied Educational Science, Umeå University. His scholarly interests are education policy, evaluation, school inspection and problems of socialisation and juridification in education. Orcid: 0000-0003-2167-6299

ORCID

Björn Högberg http://orcid.org/0000-0002-0199-0435
Joakim Lindgren http://orcid.org/0000-0003-2167-6299

References

Apple, M. W. 2005. “Education, Markets, and an Audit Culture.” *Critical Quarterly* 47 (1–2): 11–29. doi:10.1111/j.0011-1562.2005.00611.x.

Auld, E., and P. Morris. 2014. “Comparative Education, the ‘New Paradigm’ and Policy Borrowing: Constructing Knowledge for Educational Reform.” *Comparative Education* 50 (2): 129–155. doi:10.1080/03050068.2013.826497.

Beer, D. 2016. *Metric Power*. Basingstoke: Palgrave Macmillan.

Biesta, G. J. J. 2004. “Education, Accountability, and the Ethical Demand: Can the Democratic Potential of Accountability Be Regained?” *Educational Theory* 54 (3): 233–250. doi:10.1111/j.0013-2004.2004.00017.x.

Burns, T., and F. Köster. 2016. *Governing Education in a Complex World*. Paris: OECD Publishing.

Castles, F. G., and H. Obinger. 2008. “Worlds, Families, Regimes: Country Clusters in European and OECD Area Public Policy.” *West European Politics* 31 (1–2): 321–344. doi:10.1080/01402380701835140.

Charlton, B. G. 2002. “Audit, Accountability, Quality and All that: The Growth of Managerial Technologies in UK Universities.” In *Education! Education! Education!: Managerial Ethics and the Law of Unintended Consequences*, edited by S. Prickett and P. Erskine-Hill, 13–28. Imprint Academic: Thorverton, UK.

Ehren, M. C. M., J. E. Gustafsson, H. Altrichter, G. Skedsmo, D. Kemethofer, and S. G. Huber. 2015. “Comparing Effects and Side Effects of Different School Inspection Systems Across Europe.” *Comparative Education* 51 (3): 375–400. doi:10.1080/03050068.2015.1045769.

Ehren, M. C. M., and J. Perryman. 2018. “Accountability of School Networks: Who is Accountable to Whom and for What?” *Educational Management Administration and Leadership* 46 (6): 942–959. doi:10.1177/1741143217717272.

Elliott, J., L. Stankov, L. Lee, and J. F. Beckmann. 2019. “What did PISA and TIMSS Ever do for Us?: The Potential of Large Scale Datasets for Understanding and Improving Educational Practice.” *Comparative Education* 55 (1): 133–155. doi:10.1080/03050068.2018.1545386.

Erkkilä, T., and O. Piironen. 2014. “Shifting Fundaments of European Higher Education Governance: Competition, Ranking, Autonomy and Accountability.” *Comparative Education* 50 (2): 177–191. doi:10.1080/03050068.2013.807643.

Esping-Andersen, G. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge: Polity.

Grek, S. 2009. “Governing by Numbers: The PISA ‘Effect’ in Europe.” *Journal of Education Policy* 24 (1): 23–37. doi:10.1080/02680930802412669.

Grek, S., M. Lawn, J. Ozga, and C. Segerholm. 2013. “Governing by Inspection? European Inspectorates and the Creation of a European Education Policy Space.” *Comparative Education* 49 (4): 486–502. doi:10.1080/03050068.2013.787697.

Harris, D. N., and C. D. Herrington. 2006. “Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half-Century.” *American Journal of Education* 112 (2): 209–238. doi:10.1086/498995.

Holloway, J., and J. Brass. 2018. “Making Accountable Teachers: The Terrors and Pleasures of Performativity.” *Journal of Education Policy* 33 (3): 361–382. doi:10.1080/02680939.2017.1372636.

Johansson, S., and R. Strietholt. 2019. “Globalised Student Achievement? A Longitudinal and Cross-Country Analysis of Convergence in Mathematics Performance.” *Comparative Education* 55 (4): 536–556. doi:10.1080/03050068.2019.1657711.

Kamens, D. H., and C. L. McNeely. 2010. "Globalization and the Growth of International Educational Testing and National Assessment." *Comparative Education Review* 54 (1): 5–25. doi:10.1086/648471.

Lewis, S. 2017. "Policy, Philanthropy and Profit: the OECD's PISA for Schools and New Modes of Heterarchical Educational Governance." *Comparative Education* 53 (4): 518–537. doi:10.1080/03050068.2017.1327246.

Lingard, B., and S. Lewis. 2016. "Globalization of the Anglo-American Approach to Top-Down, Test-Based Educational Accountability." In *the Handbook of Human and Social Conditions in Assessment*, edited by G. T. L. Brown, and L. R. Harris, 387–403. Abingdon: Routledge.

Lingard, B., W. Martino, and G. Rezai-Rashti. 2013. "Testing Regimes, Accountabilities and Education Policy: Commensurate Global and National Developments." *Journal of Education Policy* 28 (5): 539–556. doi:10.1080/02680939.2013.820042.

Lundahl, C., M. Hultén, and S. Tveit. 2017. "The Power of Teacherassigned Grades in Outcome-Based Education." *Nordic Journal of Studies in Educational Policy* 3 (1): 56–66. doi:10.1080/20020317.2017.1317229.

Maroy, C., and X. Pons. 2019. *Accountability Policies in Education. A Comparative and Multilevel Analysis in France and Quebec*. Springer International Publishing.

Meyer, H.-D., and A. Benavot. 2013. "Introduction: PISA and the Globalization of Education Governance: Some Puzzles and Problems." In *PISA, Power, and Policy: The Emergence of Global Educational Governance*, edited by H.-D. Meyer, and A. Benavot, 9–26. Oxford: Symposium Books.

Mulgan, R. 2000. "'Accountability': An Ever-Expanding Concept?" *Public Administration* 78 (3): 555–573. doi:10.1111/1467-9299.00218.

Müller, J., and F. Hernández. 2010. "On the Geography of Accountability: Comparative Analysis of Teachers' Experiences Across Seven European Countries." *Journal of Educational Change* 11 (4): 307–322. doi:10.1007/s10833-009-9126-x.

OECD. 2016. *PISA 2015 Results (Volume I): Excellence and Equity in Education, PISA*. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264266490-en.

OECD. 2017. *PISA 2015 Technical Report*. Paris: OECD Publishing.

Ozga, J. 2011. "Governing Narratives: "Local" Meanings and Globalising Education Policy." *Education Inquiry* 2 (2): 305–318. doi:10.3402/edui.v2i2.21982.

Ragin, C. C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

Ramirez, F. O. 2012. "The World Society Perspective: Concepts, Assumptions, and Strategies." *Comparative Education* 48 (4): 423–439. doi:10.1080/03050068.2012.693374.

Ranson, S. 2003. "Public Accountability in the Age of Neo-Liberalism." *Journal of Education Policy* 18 (5): 459–480.

Ranson, S. 2008. "The Changing Governance of Education." *Educational Management Administration & Leadership* 36 (2): 201–219. doi:10.1177/1741143207087773.

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.

Schriewer, J. 2006. "Comparative Social Science: Characteristic Problems and Changing Problem Solutions." *Comparative Education* 42 (3): 299–336. doi:10.1080/03050060601022640.

Schriewer, J. 2012. "EDITORIAL: Meaning Constellations in the World Society." *Comparative Education* 48 (4): 411–422. doi:10.1080/03050068.2012.737233.

Schweisfurth, M., and J. Elliott. 2019. "When 'Best Practice' Meets the Pedagogical Nexus: Recontextualisation, Reframing and Resilience." *Comparative Education* 55 (1): 1–8. doi:10.1080/03050068.2018.1544801.

Sobe, N. W. 2015. "All that is Global is not World Culture: Accountability Systems and Educational Apparatuses." *Globalisation, Societies and Education* 13 (1): 135–148. doi:10.1080/14767724.2014.967501.

Troman, G. 1989. "Testing Tensions: the Politics of Educational Assessment." *British Educational Research Journal* 15 (3): 279–295. doi:10.1080/0141192890150305.

Verger, A., M. Moschetti, and C. Fontdevila. 2020. "How and Why Policy Design Matters: Understanding the Diverging Effects of Public-Private Partnerships in Education." *Comparative Education* 56 (2): 278–303. doi:10.1080/03050068.2020.1744239.

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. doi:10.1080/00131911.2019.1522045.

Willis, D. 1992. "Educational Assessment and Accountability: A New Zealand Case Study." *Journal of Education Policy* 7 (2): 205–221. doi:10.1080/0268093920070206.

Xiaomin, L., and E. Auld. 2020. "A Historical Perspective on the OECD's 'Humanitarian Turn': PISA for Development and the Learning Framework 2030." *Comparative Education* 56: 503–521. doi:10.1080/03050068.2020.1781397.

You, Y. 2017. "Comparing School Accountability in England and its East Asian Sources of 'Borrowing'." *Comparative Education* 53 (2): 224–244. doi:10.1080/03050068.2017.1294652.

Zapp, M., and F. O. Ramirez. 2019. "Beyond Internationalisation and Isomorphism – The Construction of a Global Higher Education Regime." *Comparative Education* 55 (4): 473–493. doi:10.1080/03050068.2019.1638103.