

Not everything that moves must converge: evidence from global policy and practice on performance-based accountability

Antoni Verger, Gerard Ferrer-Esteban and Clara Fontdevila

Introduction

Policy convergence is an often-assumed outcome of transnational policy movement. The policy convergence thesis can be summarized in that, as a result of globalization pressures and the increasing role of international organizations in policy activities, systems tend to develop “similarities in structures, processes and performance” (Bennett, 1991, p 215). Yet, in recent years, different scholars have drawn attention to the limitations of conventional approaches to convergence and to the need to unpack and critically interrogate the assumptions that inform this line of inquiry. Some scholars have thus advanced towards a multidimensional understanding of convergence, going beyond the policy adoption stage and paying greater attention to local implementation and enactment dynamics. Others have turned to the study of divergence patterns in an attempt to identify those points of mediation that explain different responses to common pressures. It is thus possible to document a shift in emphasis, from a focus on convergence patterns to an emphasis on policy variation and its causes.

The global spread of performance-based accountability (PBA) offers an opportunity to engage in these debates in an empirically informed manner. Indeed, PBA has acquired so-called *global status* (see Steiner-Khamsi, 2004) in education reform agendas, with most middle- and high-income countries adopting national large-scale assessments with the purpose of measuring academic performance and making schools more accountable (Sahlberg, 2016). However, there is limited clarity as to whether the expansion of PBA can truly be equated to the advancement of a “world testing culture” or to the transition into a single, universal accountability regime. Different scholars have drawn attention to the heterogeneity of PBA regimes and practices in place, and to the uneven level of penetration of national assessments and accountability instruments in the daily life of

schools. Nonetheless, the sources of variation behind such heterogeneity remain largely understudied.

In this chapter, we aim to overcome some of these limitations by interrogating the convergence thesis in relation to the policy implementation stage, based on a deductive design oriented at assessing the impact of theoretically plausible sources of variation. We depart from the premise that, despite the international expansion of PBA, its actual implementation in schools varies significantly across different countries. We anticipate that a range of administrative and political factors contribute to the variation of PBA practice internationally, serving as intermediary influences. To develop these arguments, this chapter is organized as follows. In the first section, we theoretically articulate the convergence versus divergence debate, and briefly review the literature on the global spread and convergence of PBA. Next, we present our conceptual framework on the points of mediation of global policy trends (in the second section) and briefly present our methods (in the third section). In the fourth section, we present our main results, and we conclude with a section discussing our findings in relation to global education policy theories and pointing to future lines of research.

The convergence thesis revisited: the case of performance-based accountability

Problematizing policy convergence

Policy convergence is generally recognized as a useful construct that brings to the fore the interdependence across educational systems, and the fact that they are often subject to common pressures and/or informed by globally circulating policy ideas. However, over the last few years, the limitations of the convergence hypothesis have come into sharp relief. Many of the limitations of convergence research stem directly from a binary understanding of policy making. Analyses of convergence trends typically focus on formalized and tangible decisions such as changes in regulatory frameworks (for example, embracing compulsory education), curricular reforms (for example, the inclusion of global citizenship education in national curriculums) or the adoption of new policy instruments (for example, national assessments). By conceiving policy making in a binary way – that is, policy being either adopted or rejected, picked or blocked (Lendvai and Stubbs, 2009) – for transfer and dissemination scholars it is relatively straightforward to determine whether educational systems are evolving to become more alike or not.

Yet, in recent years, different scholars have drawn attention to the need to go beyond the adoption stage and consider manifestations of policy other than the policy texts in which public policy is formalized. This is the case

of Christopher Pollitt, who critically notes that most transfer scholars focus on the analysis of so-called *decisional convergence*, which happens when public authorities formally “decide to adopt a particular organizational form or technique” (Pollitt, 2002, p 477). Nonetheless, to him, beyond formal policy decisions, there are many other dimensions in the convergence equation that should be also contemplated, including *discursive convergence*, *practice convergence* and *results convergence* (Pollitt, 2002). This multidimensional understanding of policy convergence implies that, however similar two policies may appear in terms of their formalization, they are likely to differ in other aspects.

Other approaches, such as policy mobilities, have similarly raised awareness on the limitations derived from an emphasis on the adoption moment. Policy mobility scholars focus on how globalizing policy ideas move through networks, are adapted and translated into “translocal settings,” and are enacted by street-level bureaucrats in ways that often depart from initial policy intentions (Peck and Theodore, 2012; McKenzie et al, 2021). To them, the idea of policy convergence is a pure mirage since, as policies move, they inevitably transform (Peck, 2011). Hence, even when two policies are labeled in the same way, they can only be nominally equivalent.

Political economy scholars are also skeptical of the convergence thesis, although for other reasons. To Colin Hay (2004), the fact that countries are exposed to common external pressures is not necessarily driving policy convergence, since countries can experience or interpret these pressures very differently, and their responses are ultimately mediated by different institutional capacities. From this perspective, the globalization of policy is thus a contingent process in which a broad number of points of mediation intervene, including political-economic regimes, policy paradigms, domestic political mediations, implementation processes, and institutional and cultural contexts (Hay, 2004).

The international spread of performance-based accountability

PBA has expanded over recent decades. This spread has taken place in parallel to a global push for decentralization and school autonomy (Salokangas and Ainscow, 2017). Autonomy and accountability tend to be depicted as inseparable policy principles in contemporary educational reform since, presumably, governments are willing to give more autonomy to schools to the extent that the latter accept stricter monitoring, and it is assumed that accountability pressure can only lead to improvement if schools have sufficient room for *maneuver* to address underperformance.

Several studies have explored the international spread of accountability-related instruments (such as national assessments, standardized testing,

performance contracts with schools and incentives schemes) and the drivers behind it (see Kamens and McNeely, 2010; Benavot and Koseleci, 2015; Verger et al, 2019a). These studies indicate that an increasing number of countries have adopted national assessments and are using these assessments for accountability purposes. Such trends have led some scholars to identify the emergence of a “global testing culture” (Smith, 2016) and to pay more attention to the externalities and outcomes of such a global phenomenon (Ramirez et al, 2018).

The expansion of PBA has also led some researchers to examine the convergence trends associated with this policy approach. Building on Program for International Student Achievement (PISA) data, Teltemann and Jude show that the use of national assessments for the purposes of accountability has intensified in Organisation for Economic Co-operation and Development (OECD) countries in recent decades, a pattern that reveals “an increased pressure on education systems to raise quality and efficiency” (Teltemann and Jude, 2019, p 268). At the same time, and on the basis of cluster analysis, the study finds that accountability trends are not homogeneous across countries, but that education systems can be split into four different groups. The authors examine such heterogeneity in the light of welfare state typologies, although they find that the explanatory power of this typology is rather limited. Following a similar approach, and also building on PISA data, Hogberg and Lindgren explore the international diffusion of accountability across OECD countries over time, with the objective to discern whether accountability is “implemented as a single coherent regime” (2021, p 301). Despite finding evidence of cross-country convergence over time, the authors also show that accountability is not a universally dominant regime and that, in fact, it is possible to distinguish between different accountability models.

To sum up, the existing research on international accountability dynamics can be divided between those who assume that the expansion of PBA instruments equates to the advancement towards a world testing culture, and those who rather show that uneven patterns prevail. The latter is certainly the case for those studies that address the implementation and enactment of PBA policies, taking advantage of the affordances offered by PISA data. Nonetheless, such studies tend to privilege inductive approaches, and the role of mediating institutions tends to be overlooked or undertheorized. To overcome these limitations, our study turns to a deductive approach designed to analyze the explanatory power of two theoretically plausible sources of variation that, according to existing literature, might contribute to explaining the divergence patterns observed by existing literature – namely, administrative regimes and partisan politics. In the following section, we discuss the theoretical underpinnings behind this decision and outline several hypotheses in relation to both factors.

Understanding global education policy through its points of mediation

The literature discussing global accountability dynamics often recognizes that there is no such thing as full convergence. Rather, accountability reforms are frequently found to differ in the rhetoric legitimizing them, as well as in their instrumentation and local practice. While numerous factors could explain this, we focus here on the role of two factors: administrative regimes and partisan politics. These factors allow us to capture two types of explanations usually mobilized to explain nonconvergence and divergence patterns – namely, path-dependency explanations (that is, how previous policy decisions and institutions influence ongoing policy) and interpretative explanations (that is, how the perception of pressures, need for change and the repertoire of changes available differ across countries) (Peters, 2021).

Administrative regimes

Bureaucratic apparatuses have long been recognized as key loci of policy making and sources of political advice, hence shaping policy choices and their enactment (Peters, 2021). To Peters and Pierre (2012), public administration matters because it does the often-invisible job of translating legal changes into action. Similarly, Christensen and Lægreid (2007) consider that reform dynamics are not only the result of external pressures, but also of how these pressures combine with polity features and historical institutional contexts. There is thus compelling evidence on the fact that administrative regimes intervene in the reception of external pressures and will decisively shape how and to what extent globally circulating ideas, such as New Public Management (NPM), are retained, adapted and enacted (Pollitt and Bouckaert, 2017).

The role of administrative traditions is often overlooked in the study of educational reform. However, there is reason to hypothesize this as a crucial variable to explain the uneven spread and implementation of PBA reforms – most notably, the fact that this is an educational reform that directly concerns the public administration and governance of education. PBA is, in fact, crucially informed by the NPM doctrine and thus echoes broader public administration debates on how to rebalance the relationship between political control, citizen-oriented services and professional autonomy. We expect those administrative regimes where the tenets of NPM have become more deeply ingrained will be more conducive to, and compatible with, the enforcement of PBA instruments in education.

To further elaborate on this hypothesis, we rely on Peters' (2021) influential four-type classification of administrative regimes – namely, Napoleonic, Germanic, Nordic and Anglo-American – which we complement, given

the sample of countries in our study, with a fifth non-Western model: the Confucian tradition¹ (see Table A1 in Appendix 1). These administrative regimes vary greatly in terms of how compatible they are with certain public sector reform approaches, and are hence key to understanding the retention and trajectory of PBA policies. At one end of the spectrum, we find Anglo-American bureaucracies. These are countries much more influenced by a deep audit culture which in fact boomed with the NPM reforms of the 1980s. Accordingly, these countries give greater priority to managerial aspects of service delivery such as performance and effectiveness, therefore facilitating the advancement of PBA.

Conversely, the centralization and uniformity that characterize civil service in Central and Southern Europe, but also in Confucian countries, is likely to hinder the implementation of certain PBA policies favoring greater managerial discretion. Likewise, the importance given to *ex ante* controls and rule compliance in many Southern and Continental European countries, or the combination of hierarchical forms of management and personal systems of loyalty that characterizes the Confucian tradition, are also likely to make it difficult to accommodate PBA postulates such as those compensating actors on the basis of their performance.

In an inbetween position, Nordic European countries appear more likely to absorb NPM postulates while adopting a selective or discriminating approach. Although these countries exhibit a greater orientation towards results achievement and are increasingly oriented to citizen demands, they have often rejected the market emphasis of the NPM agenda (including the overall idea that competition between providers will lead to performance improvements) and retain the notion that the state is ultimately responsible for addressing social problems (Pollitt and Bouckaert, 2017).

The emphasis of public administration theory in categorizing countries does not entail a static view of administrative traditions. Although administrative traditions, as historic institutions, have an important element of path-dependence, these traditions are constantly evolving. Importantly for the purposes of this chapter, numerous countries have made significant adjustments to their public administration as a result of the ascendancy enjoyed by NPM and public choice in recent decades (Christensen and Lægrid, 2007). The advent of NPM has been particularly consequential for Anglo-American and Nordic regimes, although the literature suggests that Napoleonic, Germanic and Confucian countries have not been fully insulated from such trends either.

Partisan politics

The significance of partisan ideologies and interests in the design of public policies is widely recognized in the field of comparative welfare state research

and comparative policy. In the realm of education in particular, factors such as the balance of power between different political forces and/or politically active organizations such as teacher unions are routinely highlighted as mediating the uneven spread of certain educational policies, and constitute the object of a vibrant body of research.

The work of Marius R. Busemeyer features among the most prominent exponents of this line of inquiry. According to Busemeyer, partisan politics hold great explanatory power as a source of variation in the configuration of education and training systems. The standard model of partisan theory departs from the premise that political parties vary in their preferred policies, and that such variations are largely a function of the interests of their electoral constituencies (Busemeyer, 2014). Broadly speaking, it is assumed that left-wing parties are more supportive than their right-wing counterparts of redistributive policies, educational expansion policies and a stronger role for the state in educational provision (Busemeyer, 2014; see also Busemeyer et al, 2020). At the same time, there is increasing evidence that the correspondence between party ideology, policy preferences and policy output is contingent and dynamic rather than obeying a static and linear pattern, and the link has only become more blurred in recent decades. This owes much to the changing nature of the electoral basis of different political parties, the emergence of new political cleavages and increasingly complex patterns of party competition (Hauserman et al, 2013). Political parties at the two ends of the political spectrum have indeed experienced a process of ideological transformation that accelerated during the 1990s with the advent of the “Third Way” (Giddens, 1998).

As a result of these changes, the impact of partisan preferences on educational policy outputs has acquired an element of unpredictability. Pro-market and education privatization reforms constitute a particularly illustrative example of these dynamics. Hence, a number of education reforms oriented at increasing school choice and competition have been launched or supported by left-wing parties. This phenomenon has been the object of different scholarly works (see, for instance, Wiborg [2015] or Gingrich [2011]), which have made it clear that the connection between partisan politics and policy outcomes is an ever-evolving and context-sensitive one.

When it comes to PBA, the connection between party ideology and policy preferences is certainly far from clear-cut. This is so because, as we have discussed elsewhere (Verger et al, 2019b), PBA instruments have a malleable quality in the sense that they can serve a wide array of policy goals (including transparency, efficiency, quality and equity, to name but a few). Consequently, PBA can accommodate a variety of political agendas, to the point that even governments with very different ideological orientations can frame this reform agenda as an appropriate means to realize their own

goals. Additionally, PBA instruments are frequently presented as “technical” or ideologically neutral and are particularly likely to operate as an area of consensus.

While there is no mechanical relationship between the ruling party and the adoption of PBA policies, the political orientation of the party in power might be more consequential for the policy design, evolution and implementation of such reforms (cf. Zehavi, 2012). For instance, in relation to autonomy reforms, Christ and Dobbins (2016) observe that because left- and right-wing forces adopt such policies with very different rationales, they end up advancing contrasting models of autonomy. Additionally, political parties in control of the government for long periods of time are in a privileged position to ensure that PBA policies unfold according to their political values – prioritizing the implementation of certain policy features, and hence shaping the meaning and enactment of such reforms. It is possible to hypothesize that, in countries long controlled by right-wing parties, market-like features of PBA schemes (such as the public posting of school performance data) are more likely to be retained over time than in countries long controlled by left-wing parties. Similarly, features such as the use of performance data to assess teachers’ effectiveness are less likely to be continued or emphasized in countries dominated by left-wing forces.

Data sources and methodological approach

In this chapter we look at how and to what extent PBA policies have advanced across different educational systems over time, and to the role of two mediation factors therein – namely, administrative regimes and partisan politics. For the reasons discussed earlier, we expect that these factors will crucially shape governance reforms in education, and the advance of PBA policies.

To this end, we conduct a series of convergence analyses of OECD’s PISA dataset spanning 55 countries and seven cycles (from 2000 to 2018). Specifically, we rely on PISA questionnaires to school principals inquiring about their daily practices.² By focusing on daily practices, we are not only able to see whether policy instruments have spread in national and subnational regulatory frameworks, but also to understand the presence and level of penetration of these instruments in schools’ daily life, routines and practices.

By tapping into PISA data, our study aims at bridging a methodological divide that has tended to make the dialogue between advocates and sceptics of the convergence thesis difficult. Indeed, these different approaches follow very different methodological strategies, with the biggest and most obvious difference being the preference for big-n studies in the case of diffusion

research and for small-n studies in policy mobilities research. Diffusion scholars usually rely on quantitative databases on policy adoption and the codification of international corpora of textbooks and legislation (see, for instance, Kim, 2019), whereas the latter focus on interviews and/or direct observation with policy actors operating at different political scales (see, for instance, the contributions in Anderson-Levitt, 2003). The emergence of international databases capturing both the adoption and the implementation of public policies, including OECD's PISA, has opened up the possibility to create a dialogue between different conceptual approaches.

To test the explanatory power of administrative cultures and partisan politics in PBA convergence trends, we rely on a deductive design. Next, we give an overview of the data sources and analytic strategy we used to operationalize our main variables. In order to group countries into administrative regimes, we relied on a classification informed by the work of different authors. Hence, while the final classification is heavily indebted to the work of Peters (2021) and Painter and Peters (2010a), we combined their work with classifications advanced by other scholars, including Pollitt and Bouckaert (2017) and Kuhlmann and Wollmann (2014). This allowed us to take into consideration the administrative traditions of some non-Western countries that also participate in PISA. This exercise gave rise to the following groups:

- NPM Marketizers: the UK, the US and Chile.
- *Rechtsstaat* Germanic: Germany and Austria.
- *Rechtsstaat* Napoleonic: France, Italy and Spain.
- Nordic/neo-Weberian: Norway, Denmark and Sweden.
- Confucian: the Republic of Korea, Singapore and Hong Kong.

To analyze the impact of partisan politics, we relied on an ad hoc classification of countries according to the ideological orientation of the prevailing political force for the period from 1995 to 2020.³ This 25-year timeframe is consistent with Bussemeyer's (2014) observations on the importance of the long-term approach to the balance of power between different partisan families. Particularly when it comes to understanding the level of implementation of politically charged policies, the prolonged control of the government on the part of a given political party is likely to have considerable explanatory power. Hence, and based on the data extracted from the *Database of Political Institutions 2020*⁴ curated by Inter-American Development Bank, we classified countries according to the following categories:⁵

- *Left-leaning countries*: countries in which the chief executive identified with the left for more than 60 per cent of the years in the period from 1995 to 2020.

- *Right-leaning countries*: countries in which the chief executive identified with the right for more than 60 per cent of the years in the period from 1995 to 2020.
- *Other*: countries in which the chief executive identified with the center for more than 60 per cent of the years in the period from 1995 to 2020, or in which neither right- nor left-wing parties have dominated for more than 60 per cent of the years in the period from 1995 to 2020.

Identification strategy

To identify convergence, we relied on a deductive design oriented at discerning convergence patterns in relation to different administrative regimes, based on *sigma-convergence* analysis. However, an analysis to see differential trends of convergence according to regimes would not make much sense if we did not first determine whether we can reject the existence of a common equilibrium. In the following discussion, we will give a more detailed overview of each one of these methods.

First, we carried out the *log t convergence test* proposed by Phillips and Sul (2007, 2009) to establish if there is overall convergence in each of the indicators of interest, that is, whether the common equilibrium is rejected. This test therefore had the objective of determining whether there is transitional heterogeneity of growth patterns across countries, which would imply transient divergence in such patterns (Phillips and Sul, 2009). The *log t test* is based on an innovative way to break down the variable of interest (Bartkowska and Riedl, 2012).

Panel data is usually decomposed as:

$$y_{i,t} = g_{i,t} + a_{i,t} \quad (1)$$

where $g_{i,t}$ embodies the systematic components, including the common factor, and $a_{i,t}$ represents transitory components. The specification that we used covers both common and idiosyncratic components in the elements $g_{i,t}$ and $a_{i,t}$. This decomposition may be represented by the transformation of (1) to the form (2) (Phillips and Sul, 2007):

$$y_{i,t} = \left(\frac{g_{i,t} + a_{i,t}}{\mu_t} \right) \mu_t = \delta_{i,t} \mu_t \text{ for all } i \text{ and } t \quad (2)$$

where μ_t is a single common component and $\delta_{i,t}$ is a time varying idiosyncratic element, which absorbs the error term and the unit-specific component. In other words, $\delta_{i,t}$ represents the transition path of each country, and μ_t is the common equilibrium growth path to all countries or economies (Méndez, 2020).

Next, we carried out a semiparametric specification to describe the dynamics of the idiosyncratic component. As Phillips and Sul (2007) did, we modeled $\delta_{i,t}$ in a semiparametric form to specify the null hypothesis:

$$\delta_{i,t} = \delta_i + \frac{\sigma_i \xi_{i,t}}{\log(t)t^\alpha} \quad (3)$$

where δ_i varies across the countries and $\xi_{i,t}$ is a time-varying function with mean 0 and variance 1. According to this specification, we can see whether all countries move to the same transition path, that is, whether cross-country convergence is reached. The null hypothesis of convergence is written as:

$$H_0 : \delta_i = \delta \text{ and } \alpha \geq 0 \quad (4)$$

As will be discussed later on, the null hypothesis was rejected by the data, indicating that multiple convergence groups of countries may exist. To test for this, we adopted a deductive approach by conducting a series of convergence tests in relation to subgroups of countries based on a predefined attribute – namely, administrative tradition. Specifically, derived from a theoretically informed classification, we tested whether different groups of countries do exhibit different convergence patterns. This method allowed us to test whether countries in the same administrative regime become more similar over time in terms of their assessment and accountability practices.

This deductive analysis relied on *sigma-convergence* tests. Whereas beta convergence is concerned by speed, sigma convergence focuses on dispersion and equates convergence to a decline in cross-country inequality or heterogeneity. That is:

$$\sigma_{t+T} < \sigma_t, \quad (5)$$

where σ_t is the time t standard deviation of $\log y_{i,t}$.

To calculate the σ -convergence (or σ -divergence if dispersion increases), we first calculated the coefficient of variation (CV) during the timespan of interest. The CV is a measure of dispersion of a frequency distribution (schools reporting that achievement data is used for various purposes) and is defined as the ratio of the standard deviation to the mean. Then, we regressed time on the CV for each year/wave:

$$CV_t = \delta_0 + \delta_1 t + u_t \quad (6)$$

Hence, σ -convergence allows us to observe the decline in the dispersion of a variable across a group of countries over time. In our case, we can say

there is sigma convergence when countries are converging as the dispersion of their level of accountability practices decreases over time.

Findings

The first step in our study was oriented at determining whether or not there is a global trend of convergence between countries in some of the most relevant indicators related to PBA. By exploring the evolution of countries in the implementation of accountability measures, as well as conducting log t convergence tests to confirm whether the data show a trend towards a common equilibrium, we found not only a great deal of heterogeneity between countries, but also that differences between countries have increased over time. Our results are therefore indicative that there is no such thing as global convergence in relation to any of the policies contemplated in this study – at least not in their actual practice. We refer to such policies as the use of standardized tests to assess students, the public dissemination of school results, the tracking of schools by an administrative authority and the making of judgments about teacher effectiveness (see Appendix 2 for a detailed account of the analysis underpinning this finding). Given this result, it is necessary to delve deeper into the dynamics of convergence according to groups of countries to confirm whether the administrative and political contexts potentially play a determining role.

Administrative regime mediations

To understand how administrative cultures intervene in PBA practice patterns, we perform a series of sigma-convergence tests on groups of countries derived from our ad hoc classification of administrative regimes. We start with an analysis of one of the most market-like uses of accountability, that is, the extent to which *schools' achievement data are posted publicly*. As our analysis shows, the implementation of this policy (measured as the percentage of schools reporting that achievement data are posted publicly) clearly varies across administrative regimes. While Figure 2.1a shows that NPM Marketizer countries are the ones more likely to publicly post student achievement data, followed by the Nordic ones, it is worth noting the limited spread of these practices within Germanic countries.

Remarkably, not all these country groups present the same level of internal homogeneity, as follows from the convergence analysis (Figures 2.1b and 2.1c). While there are few differences between countries classified as NPM Marketizers (that is, the coefficient of variation tends to be quite low), Nordic, Confucian or Napoleonic countries appear to constitute much more heterogeneous groups. The case of Germanic countries is interesting because they diverged during the 2009–2012 period, when the coefficient

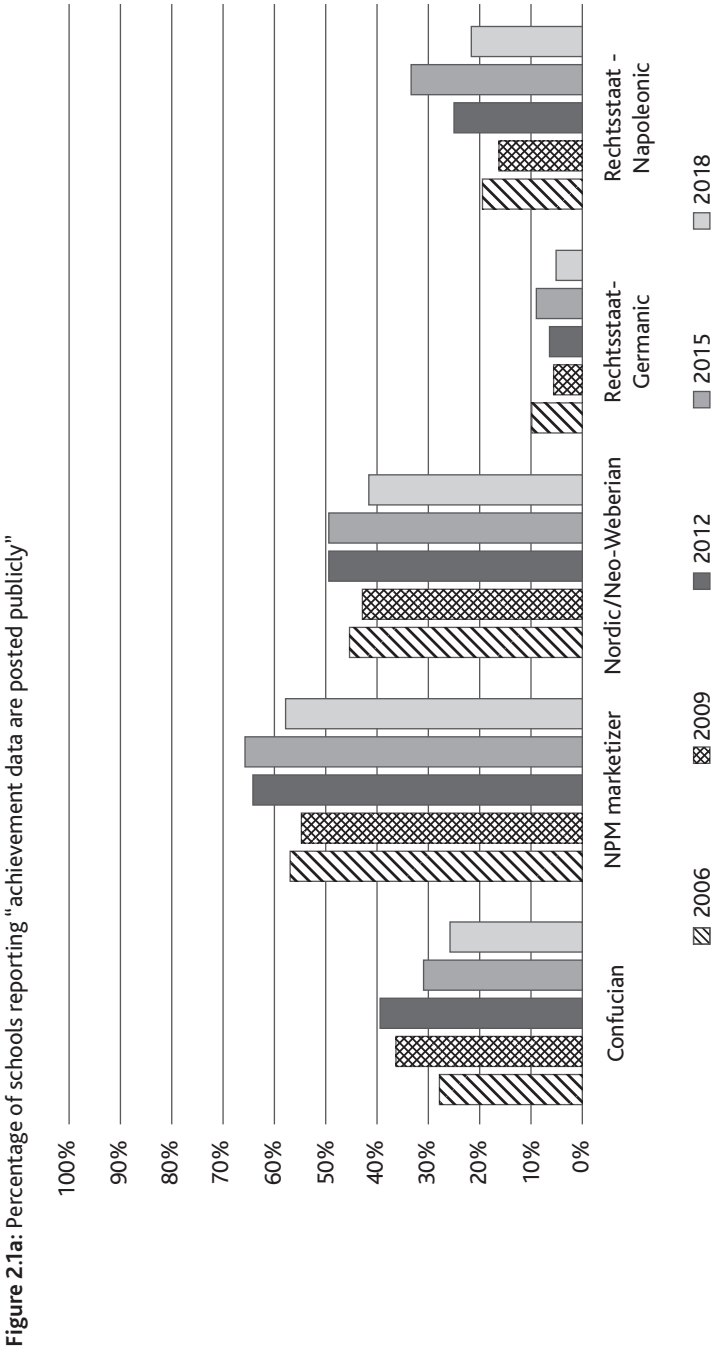


Figure 2.1b: Achievement data posted publicly: CV 2006–2018

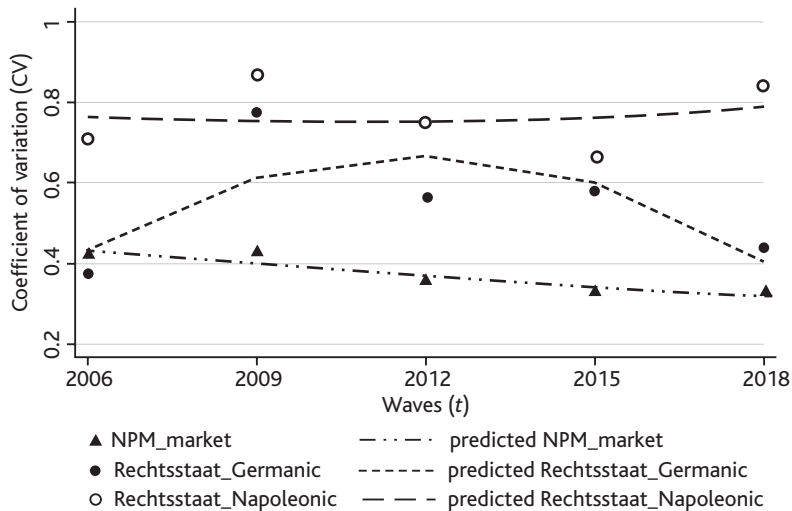
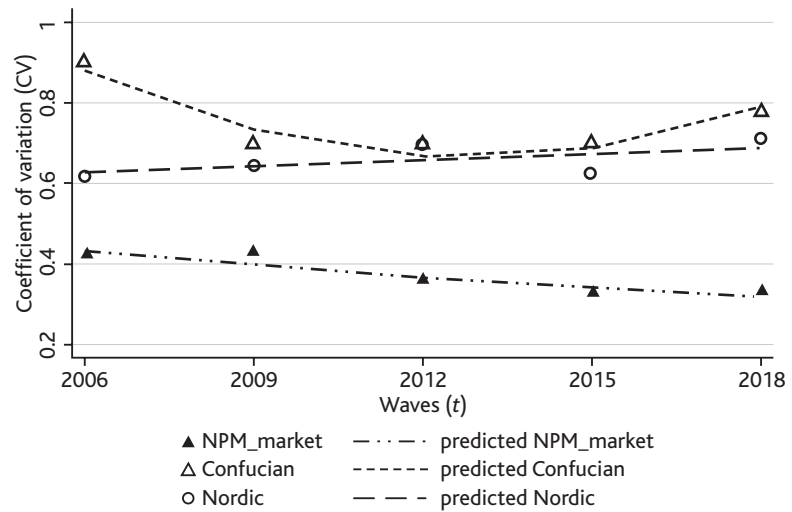


Figure 2.1c: Achievement data posted publicly: CV 2006–2018



of variation rose, and have been regaining a sense of convergence since then (the coefficient of variation has returned to the original values). Napoleonic countries, on the contrary, do not exhibit a diverging behavior, but only because they have constituted a very heterogeneous group since the very start. Confucian countries show an uneven trend towards convergence – hence, the coefficient of variation of this group experienced a decline in

the 2006–2015 period, but it has experienced an upsurge in the last PISA round included in our analysis.

Important differences are also observed in the case of schools reporting the *use of standardized tests to make judgments on teacher effectiveness* – a clear example of high-stakes accountability (see Figure 2.2a). Here, Confucian countries stand out as the ones exhibiting the greatest use of standardized tests to make judgments about teacher effectiveness. These countries also show a significant trend of convergence in relation to this practice (cf. Figure 2.2b). The second group of countries that report a significant use of test data to make judgments about teacher work are NPM Marketizers. These countries, however, show no convergence/divergence trend over time (see Figure 2.2b), which probably means that this practice has a long tradition in this type of country, with the adoption of such policies typically preceding the advent of PISA (and thus beyond the timespan covered by our dataset).

The rest of the country groups show a limited use of test data to evaluate teaching work (with less than 40 per cent of schools reporting such practices) and exhibit different convergence patterns. Napoleonic countries have tended to become more similar (especially in the 2000–2012 period), whereas Nordic and Germanic countries experimented a process of convergence in the 2000–2012 period, though their level of internal dispersion has been growing since then (see Figures 2.2b and 2.2c).

Finally, we analyze trends in the *tracking of school achievement data by an education authority*. This is the indicator that behaves more evenly across administrative regimes, at least in terms of frequencies, which means that administrative forms of accountability are those that have spread further internationally. As can be observed in Figure 2.3a, NPM Marketizer countries are again those reporting the highest level of this administrative practice, but the levels are also high in the context of the other regimes. Germanic countries are those in which achievement data are less frequently tracked by educational authorities.

In terms of convergence trends (see Figures 2.3b and 2.3c), Confucian countries appear to be the most heterogeneous group – that is, the one exhibiting higher levels of heterogeneity when it comes to tracking schools' achievement data. They do not exhibit a diverging behavior over time, but only because internal dispersion has been the rule since the start of this group regarding this practice. At the other end of the spectrum, there are the Germanic countries, which show higher levels of homogeneity: they all exhibit low levels of achievement data monitoring by the authorities. They do not converge over time, as the group is relatively homogeneous since the start of the analyzed period. As for the rest of groups, since 2012, NPM Marketizers and Napoleonic countries have tended to converge (as the slight decline of the coefficient of variation indicates), whereas Nordic

Individual use only. Not for distribution. BUP copyright material.

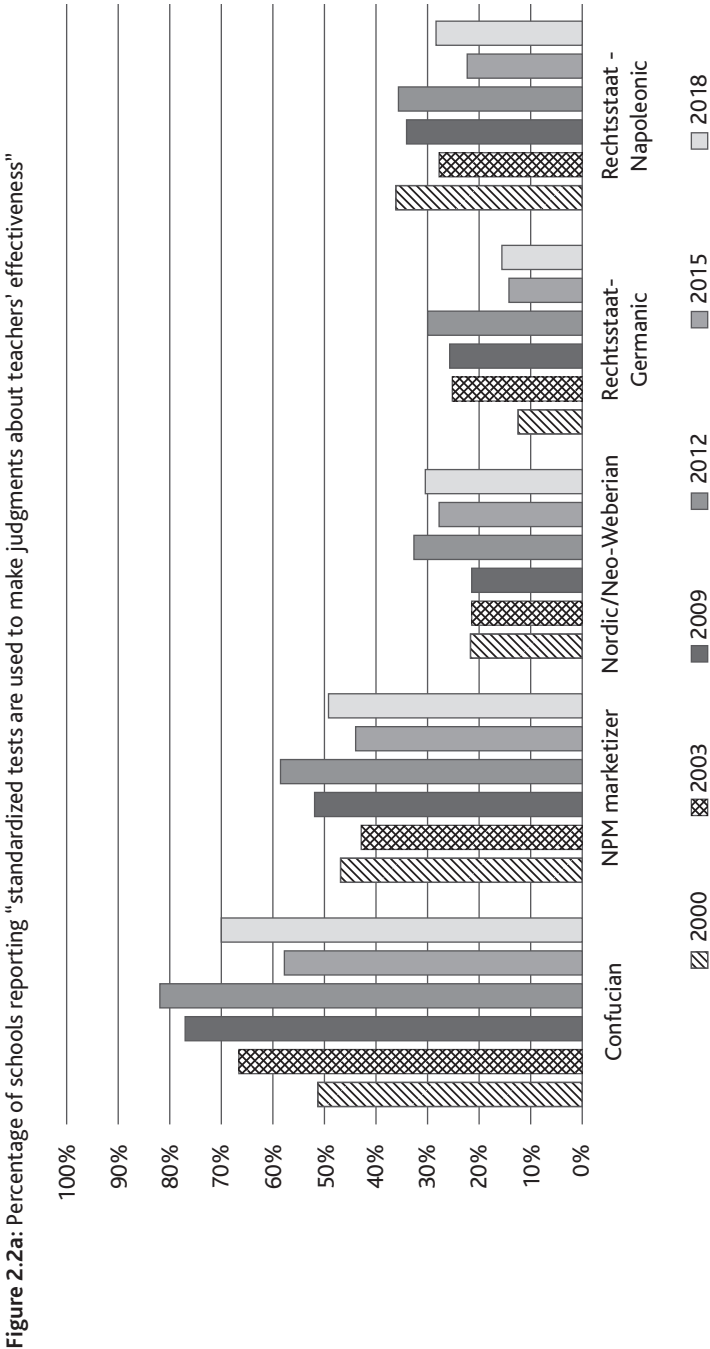


Figure 2.2b: Tests to make judgments about teachers' effectiveness: CV 2000–2018

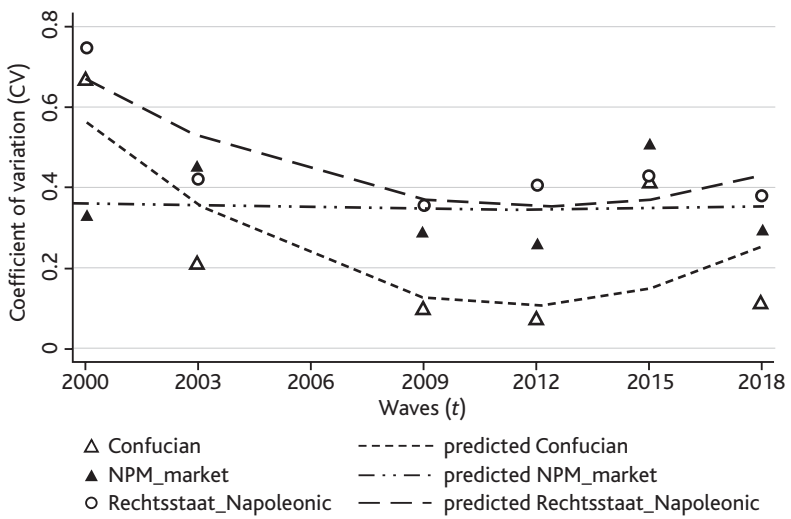
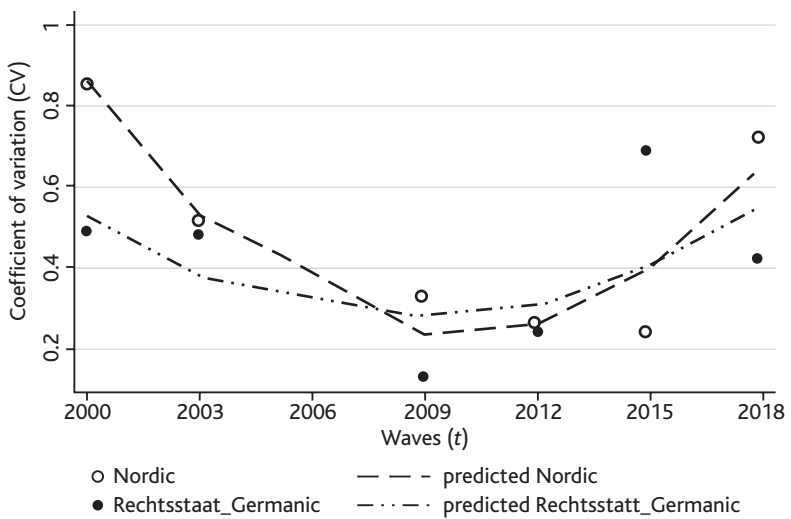


Figure 2.2c: Tests to make judgments about teachers' effectiveness: CV 2000–2018



Individual use only. Not for distribution. BUP copyright material.

Individual use only. Not for distribution. BUP copyright material.

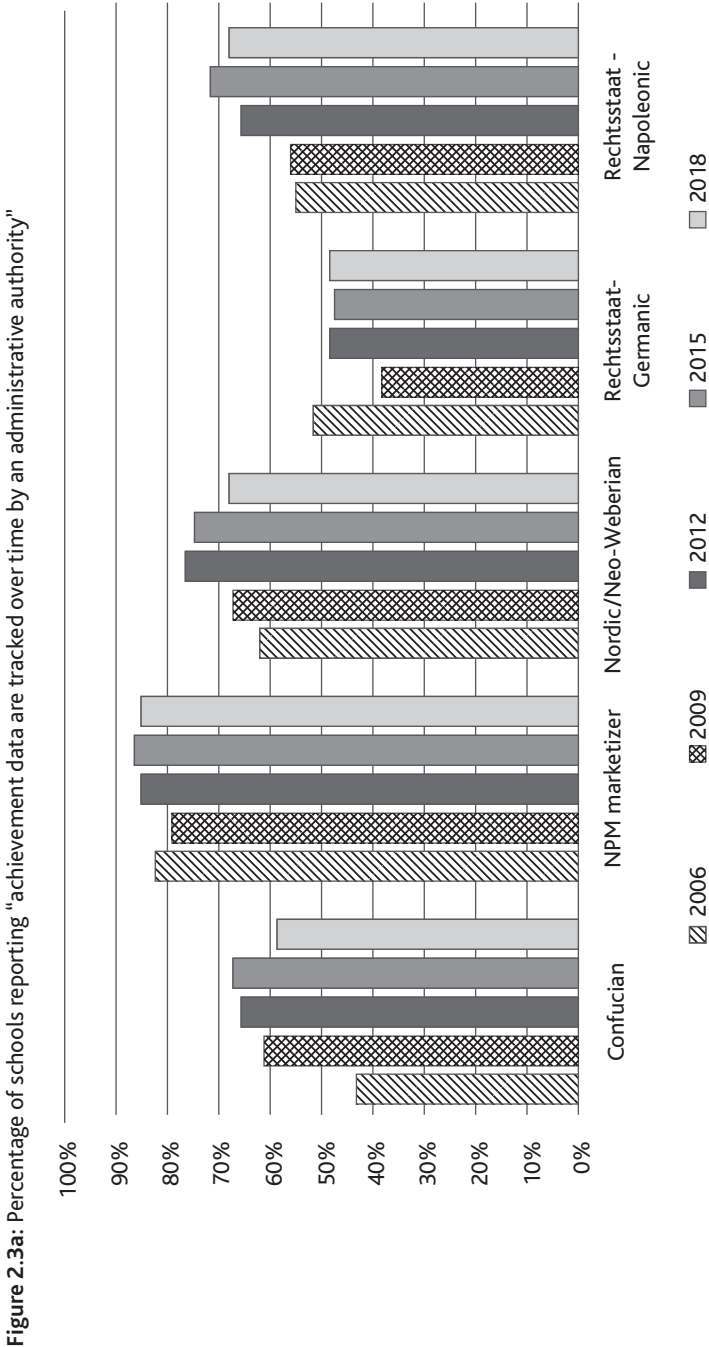


Figure 2.3b: Achievement data tracked over time by an administrative authority: CV 2006–2018

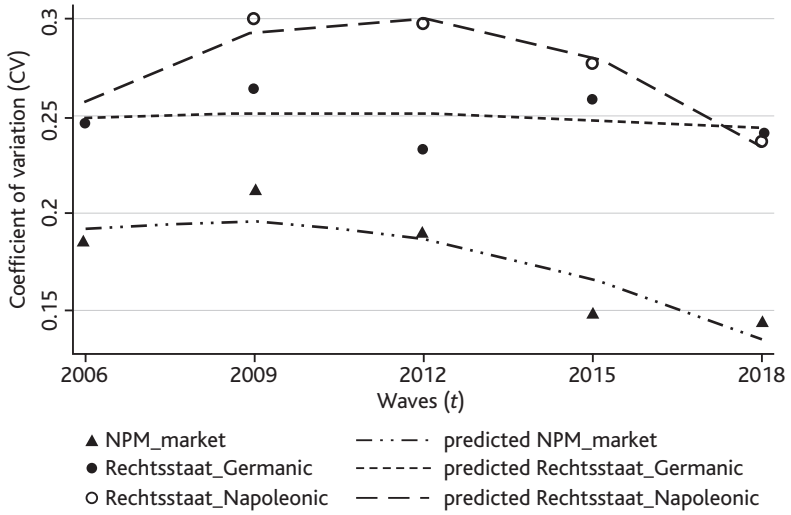
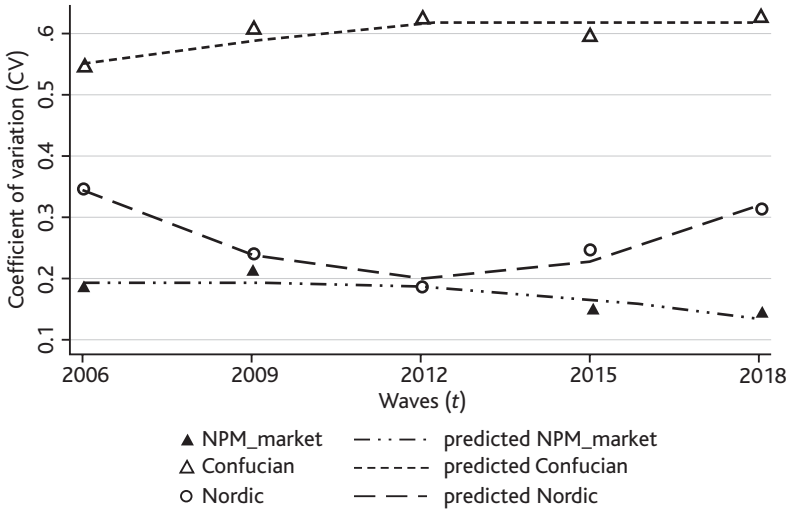


Figure 2.3c: Achievement data tracked over time by an administrative authority: CV 2006–2018



Individual use only. Not for distribution. BUP copyright material.

countries have tended to diverge (as the slight increase of the variation coefficient indicates).

Political mediations

Next, we examine how partisan politics might contribute to shaping PBA implementation patterns. To this end, we perform a series of sigma-convergence tests to groups of countries coded according to the prevailing political force in the period from 1995 to 2020.

Our findings are indicative of the fact that countries where the prevailing political force over the last 25 years is left-wing show greater convergence in accountability policies, both administrative and market-oriented. Starting with the *share of schools publicly posting data* (a proxy for market-oriented accountability, as we argued earlier), the first relevant result is that the percentage of schools that report achievement data to be publicly posted does not exceed 50 per cent in the entire time period analyzed, regardless of the ideological color of the political forces. Our results show that differences between the groups of countries are modest, and it is not possible to discern a clear pattern (Figure 2.4a). However, and somewhat counter-intuitively, left-leaning countries appear to constitute the less heterogeneous group, as the decline of the coefficient of convergence indicates that they have tended to converge since the early 2010s (Figure 2.4b). A similar pattern can be observed in relation to trends in the *tracking of school achievement data by an education authority*. As shown by the figures that follow, not only are left-leaning countries more likely to engage with this practice than right-leaning countries (Figure 2.4a), but, more importantly, they also constitute the most homogeneous group and even exhibit a modest decline in the coefficient of variation during the late 2000s.

A very different picture emerges when it comes to the behavior of right-leaning countries. These appear comparatively slightly less likely to engage in both market-oriented and administrative forms of accountability (Figures 2.4a and 2.5a). However, what is relevant here are not the differences in the practice levels of such policies, which are negligible, but rather the greater heterogeneity among right-wing countries, which indicates a more erratic implementation process and/or a reform agenda that is more contingent and permeable to the local context. Countries with predominantly right-wing governments already started from a higher coefficient of variation in 2006, thus showing greater heterogeneity between countries with the same prevailing political orientation (from 1995 to 2020), both in terms of publicly posting the achievements (Figure 2.4b) and in tracking the schools' achievement data (Figure 2.5b). In addition, the fact that these countries show a further net increase in the dispersion in both variables between 2006 and 2018, added to the fact that left-oriented countries observe a decrease

Figure 2.4a: Percentage of schools reporting “achievement data are posted publicly”

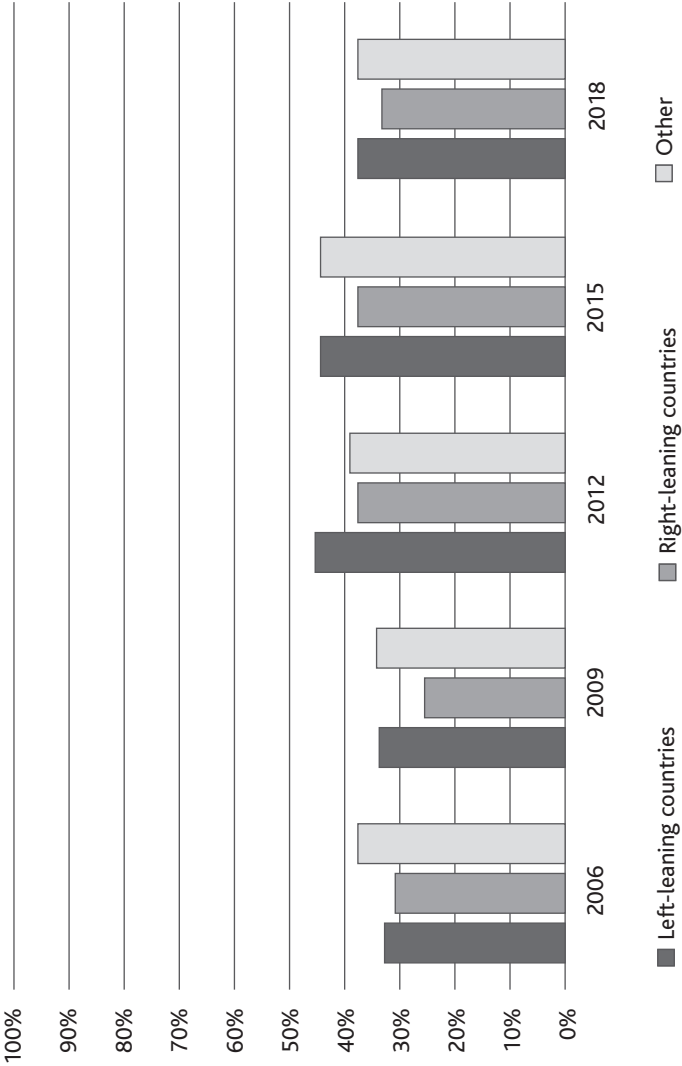
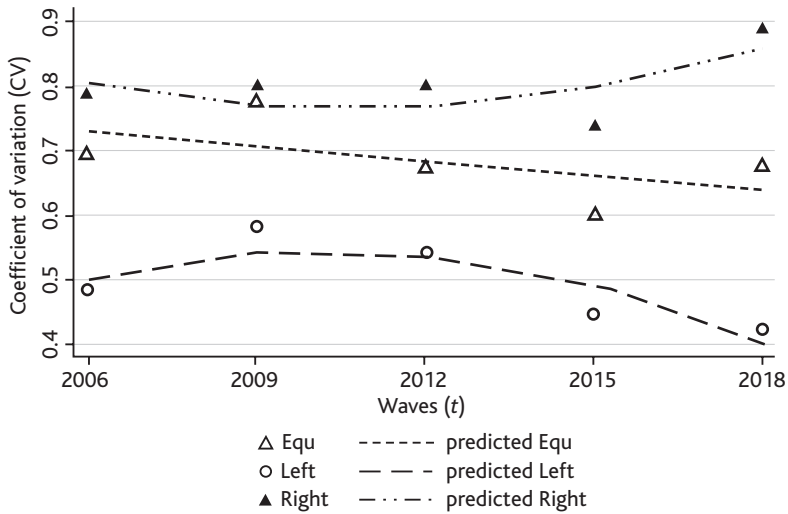


Figure 2.4b: Achievement data posted publicly: CV 2006–2018

in such variation (Figure 2.4b) or remain in a steady state (Figure 2.5b), shows that the differences between both groups of countries have widened.

Finally, we observe that partisan politics seem to have some explanatory power in the use of tests to make judgments about teachers' effectiveness, although the trend in terms of convergence does not seem to depend on the prevailing political force. First of all, even if the use of the data for this purpose barely exceeds 50 per cent of the schools in all the country groups, left-leaning countries seem comparatively less likely to engage in such practices (Figure 2.6a). Besides, there has been an overall reduction in the dispersion of values over time during the period from 2000 to 2012, meaning an increase in convergence, which is then followed by an upsurge over the last years in right-leaning countries, as well as a trend towards a steady state in left-leaning countries (Figure 2.6b). Finally, differential trends are observed between the three groups of countries if we analyze the differences between 2000 and 2018. While at the beginning of the period all the countries showed similar dispersion coefficients, showing a similar level of heterogeneity in the use of data to assess teachers, the three groups end up showing significant differences in the coefficients. As observed before, the group of left-wing countries end up having greater internal homogeneity, while the countries in the right-wing group are more heterogeneous.

Discussion and conclusions

This chapter has shed light on current trends of PBA policy and practice within education. First, it reveals – through an examination of the OECD's

Individual use only. Not for distribution. BUP copyright material.

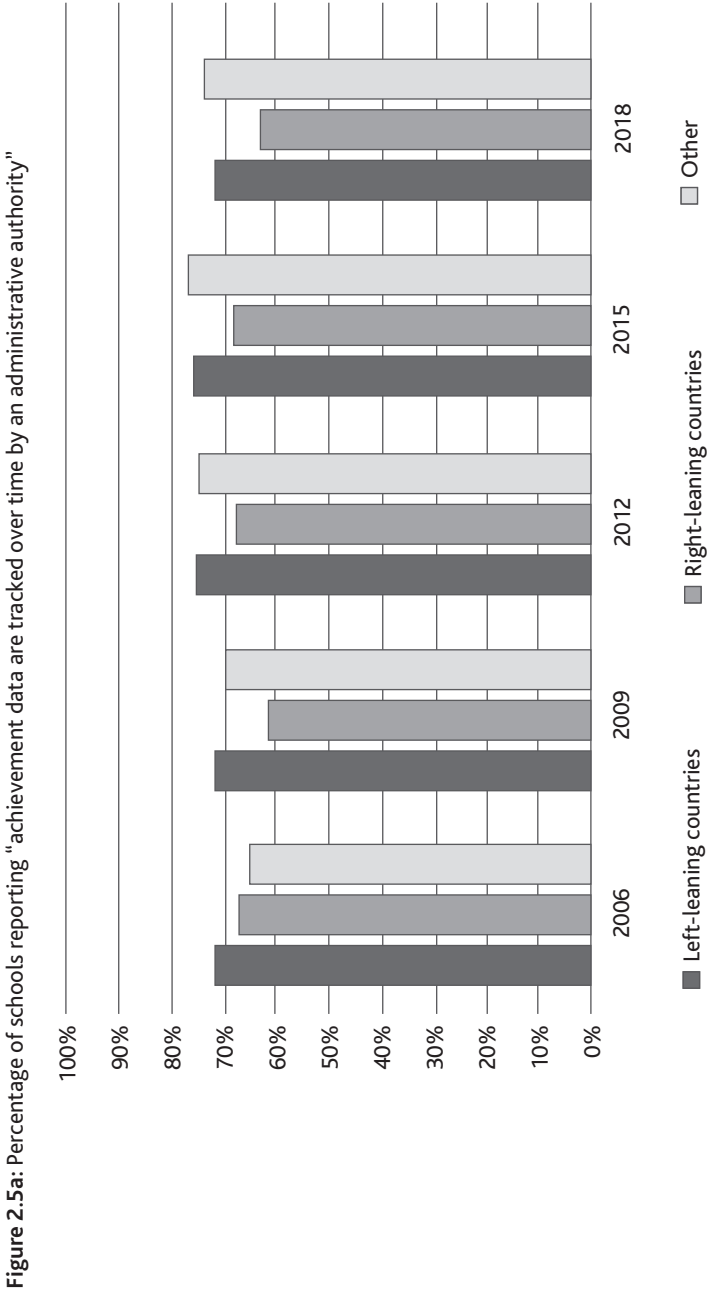
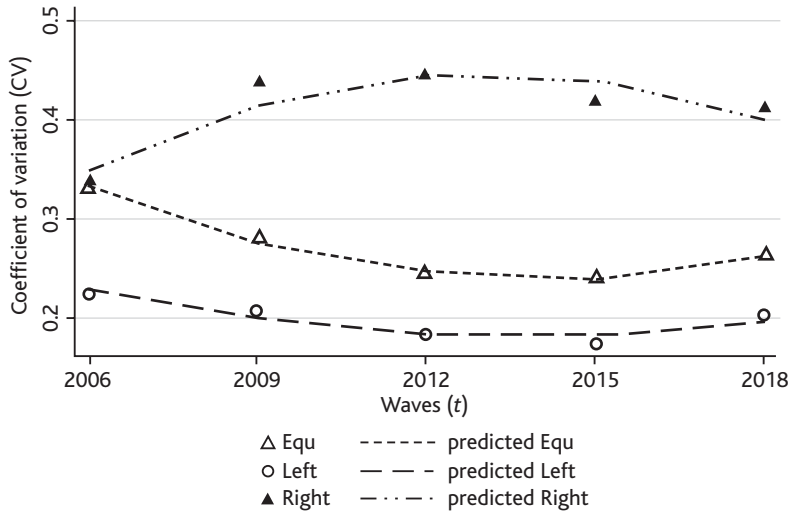


Figure 2.5b: Achievement data tracked over time by an administrative authority: CV 2006–2018



PISA dataset spanning 55 countries and seven cycles (from 2000 to 2018) – that there is no uniform pattern of converging practice regarding accountability. While it is true that many countries participate in similar policy discussions about PBA and incorporate large-scale assessments and rectifying measures for underperforming schools into their legal structures, the actual uses and degree of enactment of these policies vary greatly across countries.

Second, our study indicates that PBA implementation patterns are influenced by administrative regimes. Specifically, we find that NPM Marketizers are more inclined to publicly disclose achievement data, indicative of market-driven accountability, while countries with Confucian educational traditions utilize test data to assess teacher performance more often. Partisan politics play a less significant role, some patterns can be observed: countries with dominant right-wing politics exhibit a more diverse range of PBA practices, suggesting a patchwork implementation that is sensitive to local specificities. In these countries, the use of data for teacher evaluation varies greatly—some engage intensively with such schemes, while others do not at all. Conversely, countries with prevailing left-wing politics show more consistency, with data use for evaluating teacher effectiveness being less frequent on average.

Third, our findings reveal that certain accountability practices, such as the monitoring of data by educational authorities, cuts across different administrative regimes. This trend may stem from this practice's relatively apolitical nature, its adaptability to different rationales, and its applicability

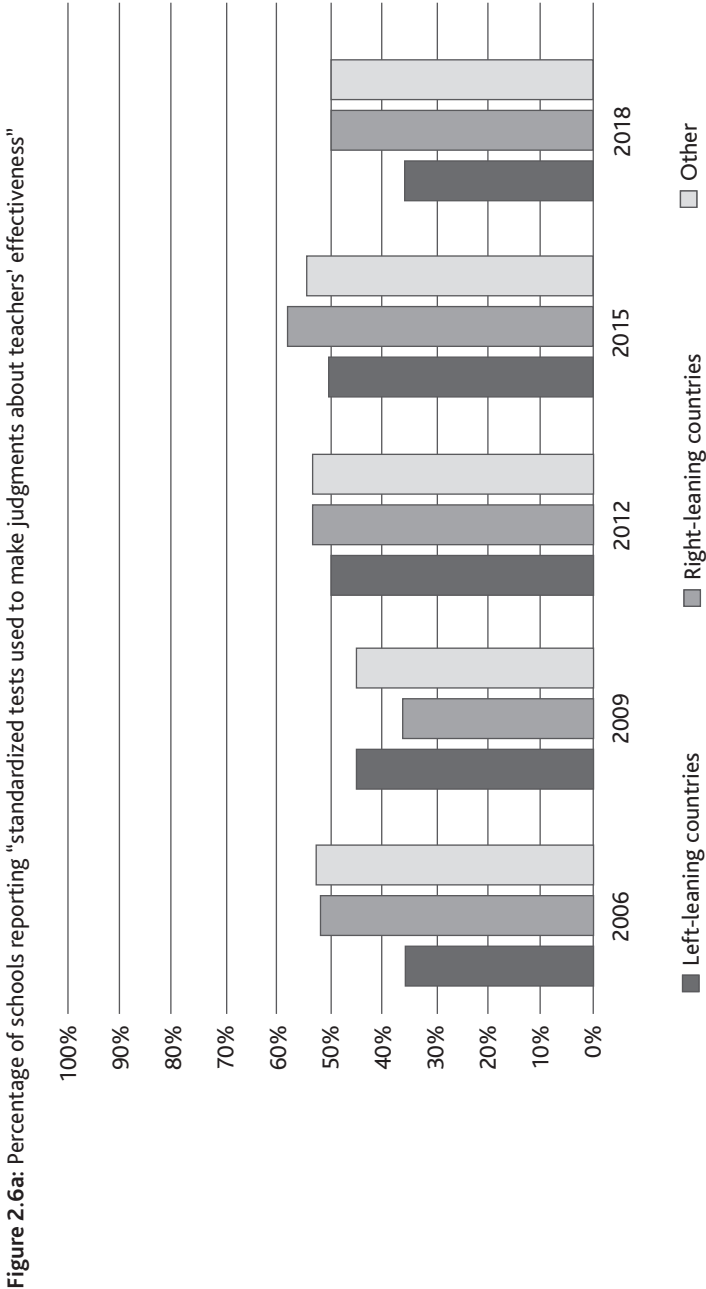
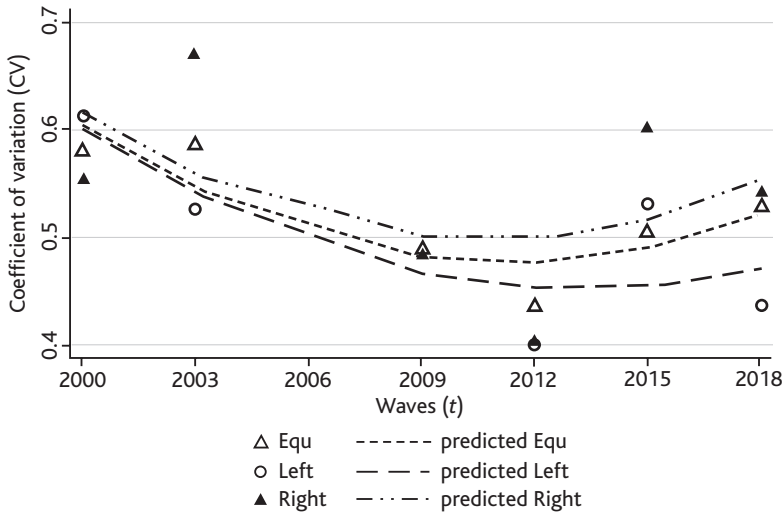


Figure 2.6b: Tests are used to make judgments about teachers' effectiveness: CV 2009–2018



within a wide spectrum of educational governance approaches. Monitoring data can serve different purposes. For instance, it acts as a tool for government oversight in education, akin to inspections, but also as a means to guide and control school autonomy within market-driven and/or decentralized environments.

Nonetheless, it is more straightforward to associate mediating institutions – especially administrative regimes – with overall implementation trends rather than with specific patterns of convergence. Within each category we analyzed from a convergence perspective, there is a notable degree of dispersion. Convergence trends are observable only in relation to certain accountability uses and country groups. For example, Confucian countries are moving towards publicizing achievement data and utilizing standardized tests to judge teachers' effectiveness. Napoleonic countries have converged in tracking school performance data more intensively, yet their convergence on other accountability measures is less consistent. The case of NPM Marketizers is intriguing; despite demonstrating low internal variation, they do not exhibit a clear convergence trend – although this is possibly because these countries implemented PBA policies well before the timeframe of our study. Moreover, grouping countries by predominant political forces does not reveal any sustained convergence in PBA practices over time.

Our research indicates that the implementation of PBA policies is often inconsistent, displaying unpredictable and, on occasion, pendular trends. Similar to other NPM instruments, PBA can lead to a host of “unintended effects, cultural surprises, discontinuities and non-linearities” (Christensen

and Laegreid, 2007, p 2). These outcomes frequently prompt a need to revise and calibrate the instruments. Therefore, the identified uneven patterns in PBA implementation must be viewed in light of well-documented implementation issues, undesirable side-effects, and complex feedback mechanisms, all of which are processes that highlight the need for more detailed investigation.

Reflections on future research

In this chapter we have focused on the role of relatively broad political and administrative categories, but future research on convergence/divergence patterns might benefit from a more fine-grained strategy. Research could pay more attention to how the ultimate impact of new governance instruments, such as PBA, depends on their intertwining with more specific features of the political architecture, including institutional rules, party politics, coalition-building strategies or the influence of instrument constituencies (cf., for instance, Busemeyer et al, 2020; Giudici et al, 2023). These elements are well-known for their role in guiding agenda-setting and policy instrumentation, yet they also critically influence how policies are put into practice. Additionally, patterns of convergence within the sphere of educational policy are also affected by factors unique to the educational sector, including the professional culture of teachers, the distribution of educational authority, and the competitive relationships among schools—topics into which we have not delved in this chapter. Therefore, the analyses we present here are intended to serve exploratory purposes, rather than leading to a definitive conclusion.

Political salience may also play a role in the policy trajectories highlighted in this chapter. The uptake of national assessments gained momentum in the 2000s as international large-scale assessments like PISA heightened competitive pressures among many countries (Fischman et al, 2019), but these uses could be losing intensity in parallel to what some observers see as a sort of PISA fatigue in recent years (Engel and Rutkowski, 2020; Jerrim, 2023), a trend further influenced by the recent COVID-19 pandemic's impact on rethinking assessment policy. The role of international competition and public opinion in the (de-)scalation of PBA pressure is yet another area ripe for further investigation.

Finally, future analyses could benefit from adopting a more transformative perspective on policy change. This would involve a nuanced examination of how global models are adopted or resisted and with what outcomes for policy systems. Reforms are often reshaped when they encounter cultural, political and institutional barriers, as well as through processes of translation and re-signification. While reforms can be either completely rejected or wholly accepted, they more frequently mutate in subtle and nuanced ways. This includes processes of policy adaptation, calibration and retooling (see, for example, Chapter 3 by Termes and Pagès, this volume) – processes that

gradually permeate existing institutions and practices. Education reform is thus a matter of degree: ideas are enacted with varying levels of intensity and focus, ultimately influencing the development of new norms, standards and practices within the educational field.

Notes

- ¹ While the Confucian style of administration is mainly associated with China, elements of this tradition have also been identified in other countries in East Asia influenced by the Chinese imperial administrative practice, including Japan, Korea, Taiwan, Vietnam and Singapore. Likewise, it is also worth noting that, in these countries, the Confucian legacy has hybridized with “imported” Western models in a complex process of institutional layering (Painter and Peters, 2010b; Drechsler, 2018).
- ² The PISA schools’ survey asks principals about their decision-making capacity, uses of large-scale assessments data, educational practices and school management styles, to name but a few areas of inquiry. Despite some limitations (most notably the fact that some policy concepts are likely to have different meanings for school principals operating in different contexts), the PISA schools database offers the possibility to understand cross-country convergence in policy decisions, but also at the level of practice.
- ³ Only countries for which data were available for at least 15 of these 25 years have been included.
- ⁴ <https://publications.iadb.org/en/database-political-institutions-2020-dpi2020>
- ⁵ Respectively labeled “Left,” “Right” and “Equ” in the figures that follow.

Appendices

The appendices in this chapter can be viewed at: <https://policy.bristoluniversitypress.co.uk/researching-global-education-policy>

References

- Anderson-Levitt, K. (ed.) (2003). *Local Meanings, Global Schooling: Anthropology and World Culture Theory*. New York: Springer.
- Bartkowska, M. and Riedl, A. (2012). Regional convergence clubs in Europe: identification and conditioning factors. *Economic Modelling*, 29(1), 22–31.
- Benavot, A. and Koseleci, N. (2015). *Seeking Quality in Education: The Growth of National Learning Assessments, 1990–2013*. Background paper for Education for All Global Monitoring Report, UNESCO.
- Bennett, C.J. (1991). What is policy convergence and what causes it? *British Journal of Political Science*, 21(2), 215–233.
- Busemeyer, M.R. (2014). *Skills and Inequality: Partisan Politics and the Political Economy of Education Reforms in Western Welfare States*. Cambridge: Cambridge University Press.
- Busemeyer, M.R., Garritzmman, J.L. and Neimanns, E. (2020). *A Loud But Noisy Signal? Public Opinion and Education Reform in Western Europe*. Cambridge: Cambridge University Press.
- Christ, C. and Dobbins, M. (2016). Increasing school autonomy in Western Europe: a comparative analysis of its causes and forms. *European Societies*, 18(4), 359–388.

- Christensen, T. and Laegrid, P. (2007). Introduction: theoretical approach and research questions. In T. Christensen and P. Laegrid (eds) *Transcending New Public Management: The Transformation of Public Sector Reforms* (pp 1–16). Farnham: Ashgate.
- Drechsler, W. (2018). Beyond the Western paradigm: Confucian public administration. In S. Bice, A. Poole and H. Sullivan (eds) *Public Policy in the “Asian Century”: Concepts, Cases and Futures* (pp 19–40). London: Palgrave Macmillan.
- Engel, L.C. and Rutkowski, D. (2020). Pay to play: what does PISA participation cost in the US? *Discourse*, 41(3), 484–496.
- Fischman, G.E., Topper, A.M., Silova, I., Goebel, J. and Holloway, J.L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy*, 34(4), 470–499.
- Giddens, A. (1998). *The Third Way: The Renewal of Social Democracy*. Cambridge: Polity Press.
- Gingrich, J. (2011). *Making Markets in the Welfare State: The Politics of Varying Market Reforms*. Cambridge: Cambridge University Press.
- Giudici, A., Gingrich, J., Chevalier, T. and Haslberger, M. (2023). Center-right parties and post-war secondary education. *Comparative Politics*, 55(2), 193–218.
- Häusermann, S., Picot, G. and Geering, D. (2013). Review article: rethinking party politics and the welfare state. *British Journal of Political Science*, 43(1), 221–240.
- Hay, C. (2004). Common trajectories, variable paces, divergent outcomes? Models of European capitalism under conditions of complex economic interdependence. *Review of International Political Economy*, 11(2), 231–262.
- Högberg, B. and Lindgren, J. (2021). Outcome-based accountability regimes in OECD countries: a global policy model? *Comparative Education*, 57(3), 301–321.
- Jerrim, J. (2023). Has peak PISA passed? An investigation of interest in international large-scale assessments across countries and over time. *European Educational Research Journal*, 55(2), 193–218.
- Kamens, D. and McNeely, C. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, 54(1), 5–25.
- Kim, Y. (2019). Global convergence or national variation? Examining national patterns of classroom instructional practices. *Globalisation, Societies and Education*, 17(3), 353–377.
- Kuhlmann, S. and Wollmann, H. (2014). *Introduction to Comparative Public Administration: Administrative Systems and Reforms in Europe*. Cheltenham: Edward Elgar.
- Lendvai, N. and Stubbs, P. (2021). Assemblages, translation, and intermediaries in South East Europe. *European Societies*, 11(5), 673–695.

- McKenzie, M., Lewis, S. and Gulson, K.N. (2021). Response: matters of (im)mobility: beyond fast conceptual and methodological readings in policy sociology. *Critical Studies in Education*, 62(3), 394–410.
- Méndez, C. (2020). *Convergence Clubs in Labor Productivity and Its Proximate Sources Evidence from Developed and Developing Countries*. New York: Springer.
- Painter, M. and Peters, B.G. (2010a). Tradition and public administration. In M. Painter and B.G. Peters (eds) *Tradition and Public Administration* (pp 19–30). Basingstoke: Palgrave Macmillan.
- Painter, M. and Peters, B.G. (2010b). Administrative traditions in comparative perspective: families, groups and hybrids. In M. Painter and B.G. Peters (eds) *Tradition and Public Administration* (pp 2–16). Basingstoke: Palgrave Macmillan.
- Peck, J. (2011). Geographies of policy: from transfer–diffusion to mobility–mutation. *Progress in Human Geography*, 35(6), 773–797.
- Peck, J. and Theodore, N. (2012). Follow the policy: a distended case approach. *Environment and Planning A: Economy and Space*, 44(1), 21–30.
- Peters, B.G. (2021). *Administrative Traditions: Understanding the Roots of Contemporary Administrative Behavior*. Oxford: Oxford University Press.
- Peters, B.G. and Pierre, J. (2012). *The SAGE Handbook of Public Administration*. London: SAGE.
- Phillips, P.C. and Sul, D. (2007). Transition modelling and econometric convergence test. *Econometrica*, 75(6), 1771–1855.
- Phillips, P.C. and Sul, D. (2009). Economic transition and growth. *Journal of Applied Econometrics*, 24(7), 1153–1185.
- Pollitt, C. (2002). Clarifying convergence: striking similarities and durable differences in public management reform. *Public Management Review*, 3(4), 471–492.
- Pollitt, C. and Bouckaert, G. (2017). *Public Management Reform: A Comparative Analysis – into the Age of Austerity*. Oxford: Oxford University Press.
- Ramírez, F.O., Schofer, E. and Meyer, J.W. (2018). International tests, national assessments, and educational development (1970–2012). *Comparative Education Review*, 62(3), 344–364.
- Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard and A. Verger (eds) *The Handbook of Global Education Policy* (pp 128–144). Oxford: Wiley-Blackwell.
- Salokangas, M. and Ainscow, M. (2017). *Inside the Autonomous School: Making Sense of a Global Educational Trend*. Abingdon: Routledge.
- Smith, W.C. (2016). An introduction to the global testing culture. In W.C. Smith (ed.) *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice* (pp 7– 23). Providence, RI: Symposium.
- Steiner-Khamsi, G. (ed.) (2004). *The Global Politics of Educational Borrowing and Lending*. New York: Teachers College Press.

- Teltemann, J. and Jude, N. (2019). Assessments and accountability in secondary education: international trends. *Research in Comparative and International Education*, 14(2), 249–271.
- Verger, A., Parcerisa, L. and Fontdevila, C. (2019a). The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, 71(1), 5–30.
- Verger, A., Fontdevila, C. and Parcerisa, L. (2019b). 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.
- Wiborg, S. (2015). Privatizing education: free school policy in Sweden and England. *Comparative Education Review*, 59(3), 473–497.
- Zehavi, A. (2012). Welfare state politics in privatization of delivery: linking program constituencies to left and right. *Comparative Political Studies*, 45(2), 194–219.