
1 Benchmarking in federal systems

Alan Fenna¹
Curtin University

One of the more notable recent developments in federal systems has been the growing use of benchmarking arrangements to improve service provision by the constituent units. Inter-jurisdictional benchmarking is in its early stages, but there is no doubting its significance. If clear indication were lacking, that changed in August 2009 with the insertion of Article 91d in the Constitution of the German Federal Republic. ‘The Federation and the States may, to establish and improve the performance of their administrations, conduct comparative studies and publicise the results.’² How much legal or practical impact such an ambiguous clause will have is unclear (see Konzendorf, this volume). What is clear, though, is its symbolic significance — benchmarking is now a recognised device of modern federalism. At the same time, the issue is far from straightforward. Only a year after Article 91d was inserted in the German Constitution, the incoming government of the United Kingdom announced that it would wind up its Audit Commission, the body that for two decades has carried primary responsibility for the extensive performance monitoring and benchmarking that has been imposed on local government in the UK (DCLG 2010; Downe 2008; Grace, this volume). The juxtaposition of these two events suggests something about the complexity of the issue: if benchmarking has been called into question in the UK where it had been instrumental in driving reform for two decades, what is its future in a federal context?

This book is about the intersection of a particular form of government and a particular tool of management. Each is a complex matter in itself. The question here is how complementary they might be. How compatible is benchmarking with principles of federalism? Under what circumstances is benchmarking likely to take hold in federal systems? To what extent can benchmarking ‘add value’ to existing federal arrangements either by offering a superior mode of intergovernmental relations and/or by generating better substantive results for citizens? In addressing

¹ Alan Fenna is a Professor of Politics at The John Curtin Institute of Public Policy, Curtin University.

² *Grundgesetz für die Bundesrepublik Deutschland*, Art 91d: *Bund und Länder können zur Feststellung und Förderung der Leistungsfähigkeit ihrer Verwaltungen Vergleichsstudien durchführen und die Ergebnisse veröffentlichen*. Inserted by amendment 1 August 2009.

those questions we are inevitably drawn into consideration of the very different forms that both federalism and benchmarking can take.

This chapter provides an overview of the problem considered as four propositions:

1. Benchmarking is a logical but challenging instrument of public sector governance that comes in different forms and carries with it a number of risks or limitations.
2. Federalism is a very specific form of government predicated on well-established norms and promising certain advantages but also one where significant differences in practice from one instance to the next make direct comparison difficult.
3. In both principle and practice, there are affinities between different types of federalism and different types of benchmarking.
4. While the experience with benchmarking in federal systems is limited and highly varied, there is a fundamental difference between those instances of a cooperative or collegial benchmarking nature and those that are ‘top-down’.

1.1 Benchmarking in the public sector

‘Benchmarking’ is a term that is used rather loosely and takes on at least two somewhat different meanings, one more demanding than the other. In the looser or broader sense we can understand benchmarking simply to mean the comparative measurement of performance. In the fuller or more specific sense we can understand benchmarking to mean the use of comparative performance measurement as a tool for identifying and adopting more efficient or effective practices (Watson 1994). In the former sense it is an assessment device; in the latter it is a learning and adjustment tool. Most generally, it ‘is not so much a technique as a way of thinking’ (OECD 1997, p. 25) — a disposition toward comparative assessment and learning. Like other aspects of the ‘new public management’, benchmarking is a practice that has spread from the private sector to the public sector with the hopes that it will drive improvements in public service delivery. It has its share of enthusiasts (Hatry 2006; Metzenbaum 2008; Osborne and Gaebler 1992; Wholey and Hatry 1992) but also sceptics (Bevan and Hood 2006; Grizzle 2002; Radin 2006; Smith 1995).

Archetypes: external and internal benchmarking

In the classic model of private sector benchmarking, an individual firm finds a way to assess performance of some aspect of its enterprise against industry leaders in other sectors and learns from that comparison how to improve its processes. The

assessed and the assessors are effectively one and the same. Such *external* benchmarking is voluntary or self-directed; the ‘audience’ is restricted to the management of the initiating firm itself; and the exercise is oriented solely toward learning.

There is another model from the private sector, though, and that is the *internal* one: central management impose benchmarking requirements on the firm’s constituent units as a disciplinary device, or way of driving improvement through internal competition. This internal benchmarking thus mimics the market forces that had been displaced by the creation of the business enterprise in the first place. In this version, the assessed and the assessors are different, and the former are subject to sanctions imposed by the latter. This is neither self-directed nor focused on learning as far as those individual units are concerned; rather, it is *top-down* and coercive, focusing on performance monitoring. Such internal performance monitoring equates to benchmarking in a broader or looser sense of the term.

From private to public sector

The evident value of performance comparison, identification of best practice, and commitment to learning and improvement — not to mention the potential to increase performance accountability — makes benchmarking an attractive proposition for the public sector as for the private and it has become an important feature of contemporary public administration (Carter, Klein and Day 1995). In Osborne and Gaebler’s (1992, p. 146) oft-cited argument, ‘what gets measured gets done’, and if public sector agencies start measuring what they do they will find ways to do it better. And if governments can shine the spotlight of performance measurement onto the things that ultimately count the most — what government *achieves* as distinct from what it merely *does* — then presumably they will find a way to achieve more.

However, while alike in some regards, the public and private sectors are distinctly un-like in some fundamental respects. One of those is that governments and their various agencies are not profit-driven enterprises engaged in a competitive struggle for business and survival in the market place. This means they are not under the same relentless compulsion to perform in objective terms. Another is that the very *raison d’être* of government is to achieve impact or *outcomes* in society rather than merely output. In that respect the tasks of government could not be more different — nor more challenging — than those of the private sector.

Thus the public sector has neither the same *imperative* nor the same *capacity* for benchmarking as the private sector. In addition, the important thing for governments

and public sector agencies is not so much to be performing, it is to be *perceived as performing* since straightforward objective assessment of governmental performance is so much more difficult and contestable. There is a single, undisputed and objective criterion of performance in the private sphere and that is profitability. Benchmarking is not done to *assess* one's performance, but to *improve* one's performance. Not so in the public sector. There, benchmarking is the assessment of performance. A correlate of this is that achieving strong performance is nowhere near as important as avoiding poor performance:

There might not be any strong incentive in performing 'best' because the 'winner' hardly 'takes it all' in public management. It may rather be that 'the loser loses it all.' For the Opposition, there is not much reward in identifying high performance. It is exposing and blaming low performance that may eventually bring the Opposition into the ministerial seats after the next election. (Johnsen 2008, pp. 172-3)

The surrogate role

Both external and internal versions of benchmarking can be found in the public sector — often referred to as, respectively, 'bottom-up' and 'top-down' benchmarking (Goddard and Mannion 2004). However, because the public sector more closely resembles a large multi-unit corporation, it is the internal, top-down, version that tends to predominate. The lack of intrinsic incentive is in some ways precisely the reason for introducing benchmarking — just as it has been for *internal* corporate benchmarking. Performance monitoring and the imposition of benchmarking requirements is a public sector surrogate for market forces. This may be initiated by an individual agency to improve its own performance — external benchmarking — but given the lower level of intrinsic incentive and the greater difficulties, such action is likely to be the exception to the rule. In reality, the lower level of incentive means that public sector agencies are more likely to need such requirements to be imposed on them.

Sanctions

Internal benchmarking operates via sanctions — which, in the private sector, appear in the form of decisions about capital allocation. In that sense, it is a coercive device. In the public sector, sanctions might take a number of forms of which two are particularly prominent. One, following the private sector lead, relies on financial penalties and rewards. There are, however, drawbacks to financial penalties — among them the distinct possibility that substandard performance may require more, not less, resource input to address. A multi-site corporation is free to let its underperforming sites wither and die; governments are not. Hence, then, the attraction of a quite different form of sanction: the political device of *naming and*

shaming. Here the exercise has the public as audience — an audience it is assumed can be reached effectively and will respond in a way that has the desired sanctioning effect. Reaching such an audience often means simplifying performance information to construct ‘league tables’ ranking jurisdictions or agencies by their performance. Well-known in the context of schools performance, this is a much-debated device (Burgess, Wilson and Worth 2010; Goldstein and Leckie 2008; Nutley and Smith 1998; Risberg 2011; West 2010).

Perverse effects

Any form of sanctioning creates incentives for behaviour contrary to the intentions of the benchmarking regime (Hood 2006; McLean, Haubrich and Gutiérrez-Romero 2007; Radnor 2008; Smith 1995). Two in particular are widely acknowledged. A focus on generating the desired results as reflected in the measurement criteria may induce ‘effort substitution’ (Kelman and Friedman 2009) such as *teaching to the test* where measured performance is enhanced by neglecting the broader suite of often less tangible or immediate desiderata. The overall purpose is eclipsed in these misguided efforts to achieve the measured targets. Since indicators are at best incomplete representations of policy objectives and sometimes vague proxies (‘synecdoche’), there is always going to be a tendency to ‘hit the target and miss the point’ (Radnor 2008). *Gaming* takes the problem one step further, with performance monitoring regimes giving agents an incentive to structure their activities in such a way as to produce the desired *indication* of results without necessarily generating any improvement in real results (‘strategic behaviour’). We could expect that the higher the stakes involved, the higher the propensity for perverse behaviour of both those forms.

It is presumably possible to design systems to address such problems (Bevan and Hood 2006). Proponents argue that good design and improvement over time will minimise pathologies and even if there are such dysfunctional responses, the overall gain may outweigh the costs (Kelman and Friedman 2009; Pollitt 1990, p. 172).

Outcomes measurement

The second of the challenges and the one that is most particular to the public sector — having a focus on outcomes rather than merely outputs — is less amenable to solution. Private enterprise judges its success by outputs; those outputs all have monetary values; and there is no debate about ultimately what the goal is. Private enterprise is not concerned with what its impact might be. Indeed, if it were, many widely available commodities and services would cease to be produced. Government produces outputs, but these outputs are only a means to an end, the end

of addressing some problem in the economy or society. The ultimate goal is *outcomes* and that presents problems of measurement, attribution and direction. *Social indicators*³ may exist or be developed for many outcomes but with varying difficulty, particularly for outcomes with longer time horizons. Schools should produce children with identifiable and testable cognitive skills; but to some degree that is an indicative or intermediate outcome. Schools ultimately should produce citizens who over the longer term prove to be capable economic agents and well-adjusted members of society. Even if the outcomes are readily measurable, they may not be so readily influenced through policy; to what factors do we attribute performance? And finally, unlike in the private sector there are legitimate differences in views about what outcomes the public sector is seeking in many areas.

Of course, there is much utility in measuring public sector outputs and in measuring output efficiency ('process benchmarking') and there are a number of practical services government provides whose 'impact' is not the issue. Even here there are not-insignificant challenges given the complexity of many public sector outputs. The argument of benchmarking advocates is that the creation of such regimes prompts and promotes progressive improvement in the data: 'a poor start is better than no start' (Osborne and Gaebler 1992, p. 156). One lesson of the UK experience with a performance monitoring reliance on quantitative indicators, though, seems to have been that significant qualitative dimensions slip through the net with potential for quite misleading conclusions to be drawn (AC 2009). For public sector benchmarking, much hinges on the development of reliable indicators, in regard to both processes and outcomes (Atkinson *et al.* 2002; Bauer 1966; DCLG 2009; Esping-Andersen 2005; Hvinden 2005; Innes 1990; Marlier *et al.* 2007; OECD 1982). In addition, it requires that data sets be fully consistent across the benchmarked entities and reasonably consistent over time. And, given the complex relationship between government action and particular economic or social desiderata and the degree to which circumstances vary, assessment of those data must be well contextualised.

False modesty?

Critics of performance management see it as being based on highly unrealistic assumptions about the availability and objectivity of information; the cause-and-

³ A social indicator has been defined as 'a statistic of direct normative interest which facilitates concise, comprehensive and balanced judgements about the condition of major aspects of a society. It is in all cases a direct measure of welfare and is subject to the interpretation that, if it changes in the "right" direction, while all other things remain equal, things have gotten better, or people better off' (DHEW, 1970, p. 97). See also Bunge (1975).

effect relationships between government actions and societal outcomes; the amenability to quantification; and the sufficiency of baseline data (Radin 2006, pp. 184–5). Unfortunately, to this point we have little hard performance data on what net benefits performance benchmarking delivers. ‘The outcomes of performance management systems are generally unmeasured and little is known about their cost effectiveness or endurance over time’ (Sanger 2008). In general, proponents of performance monitoring and benchmarking hasten to qualify their ambitions with the caveat that, as an early proponent put it over a century ago, ‘In answer to the objection that figures mislead, the obvious reply is, figures do not talk. They may raise questions; they do not answer questions’ (Allen 1907, p. 125; Kravchuk and Schack 1996). In this conception, performance data have the relatively modest role of identifying problems for analysis and assessment — raising questions rather than providing answers. However, this may be falsely modest given the propensity for performance data to be seized upon as objective evidence of success or failure.

1.2 Concerning federalism

It is, of course, not just the challenges of implementing a benchmarking regime in the public sector that we are interested in here, but the challenges of doing so between jurisdictions in a federal system. Four main features of federalism are particularly relevant. First, federalism is — most certainly in principle if to a slightly lesser degree in practice — a distinct mode of governance and not simply of matter of centralisation versus decentralisation or ‘multilevel governance’. Secondly, federalism is widely held to offer certain advantages or benefits as a system of government. Thirdly, there are a relatively small number of established federations, with a relatively high degree of variation between them — so comparison and generalisation are difficult. And fourthly, in practice federal systems have developed a high degree of operational complexity.

The distinctiveness of federalism

Federalism is a particular form of constitutionalised power sharing whereby sovereignty is in some sense and to some degree shared and powers divided between two levels of government, *viz.*, the central government and the governments of the territorially-defined constituent units (Hueglin and Fenna 2006). It is predicated on three main tenets. The first is that the two levels have a constitutionally-protected autonomy: neither level can unilaterally alter the status or roles of the other. The second is that constituent units have a meaningful degree of responsibility for local matters. And the third is that for matters affecting all,

decisions are made nationally not locally. Taken together, these last two principles are similar to the European Union's *subsidiarity* principle: the rule that tasks should be performed by the lowest level of government that can execute them effectively.

There are at least two corollaries of these defining principles. One is that the member governments of a federation are accountable first and foremost to their own political communities and not to each other or to the wider national community. It is not for the national community to punish or over-rule local communities for 'bad' policy or politics. The other is that relations between the two levels of government in a federation be conducted in accordance with principles of mutual respect.

The (putative) benefits of federalism

While federalism emerged as a practical expedient — a way of achieving the scale benefits of union without forsaking autonomy — it has come to be seen as possessing certain virtues as a mode of government. Traditionally, the first of these has been seen as being the protection of legitimate difference and the ability to have policy tailored to local needs and preferences. Scope for, and interest in, such diversity has, declined greatly over the last century, but this remains an important consideration in the case of pluri-national or pluri-lingual federations. Three other suggested advantages of federalism have been widely canvassed. The first of these is *local accountability*. The second is so-called *laboratory federalism*, viz., the enhancement of policy-learning capacity through the multiplication of policy-making sites (Bryce 1893, p. 353). And the third is *competitive federalism*, viz., the ability of citizens to compare the performance of their government with that of governments in other jurisdictions, otherwise known as 'yardstick competition' (Salmon 2006).

These are, however, theoretical or hypothetical advantages. Whether they are actually realised — or are realised to an extent that compensates adequately for the inevitable disadvantages of divided jurisdiction — is a matter for empirical assessment. Divided jurisdiction blurs lines of accountability; it is not always easy for citizens to compare performance across jurisdictions meaningfully; and jurisdictions engage in insufficient policy experimentation and have limited scope for learning from one another.

The diversity of federal systems and experiences

Actually-existing federations each have their own character. Each is distinct, and generalisation is difficult. They differ in constitutional design; in their evolution and

development over time; and in their underlying characteristics. The upshot is that one has to make highly contextualised comparisons.

This book covers examples from five major federations: Australia, Canada, Germany, Switzerland and the United States. It also includes one federation-in-the-making, the European Union, and one unitary state, the UK. We should note here that although its experience with benchmarking between levels of government has been that of a unitary rather than a federal state, the UK example is of direct relevance simply because of its extensive and pioneering record in the area (see Grace, this volume). Indeed, with this year's dismantling of the Audit Commission, we can see the UK experience as one of the very few that has run its full course.

Divided versus integrated federalism

The three Anglo federations share some important commonalities. Most importantly, they are all based around the legislative division of powers with the two levels, at least ostensibly, exercising full powers of policy-making, implementation and administration within their assigned spheres (Hueglin and Fenna 2006). This sets them apart from the German model, which assigns policy-making responsibility in many areas to the central government and responsibility for implementation and administration to the constituent units, the *Länder* (Kramer 2005; Schneider 2006). As a corollary of that approach, the German model incorporates a heightened degree of representation — via the upper house of parliament, the *Bundesrat*, or Federal Council — for the constituent units in the legislative process of the central government (Oeter 2006).

Differences within differences

On paper, the similarity between the Australian and American federations is particularly strong given the degree to which the Australian founders followed the American example in their design and drafting (Saunders 2005). However, even among the Anglo federations differences are significant. Of relevance are the fact that the US uses a presidential 'separation-of-powers' form of government while Canada and Australia are parliamentary systems; the much larger number of units in the American system; the distinctiveness of Canada as a federation divided between an English-speaking majority and a French-speaking minority centred in one of the main provinces; and the absence of any significant underlying federal difference in Australia.

Degree of centralisation

Federations range from the quite decentralised Swiss and Canadian cases to the highly centralised Australian one. Although Switzerland leans strongly toward the German model of administrative federalism, it does so in a much less centralised or regionally uniform way (Braun 2010, pp. 169–70). With its strong cantonal and local identities, long history of confederalism, different language communities, and unique reliance on direct democracy, Switzerland is a federation with unusually strong federal characteristics (Armingeon 2000; Linder 2010; Schmitt 2005). Nonetheless, even Switzerland feels the centralising pressures endemic in federal systems. Meanwhile, Australia and the United States have developed highly centralised characteristics in a number of areas. In both countries, for instance, conditional grants have been used for many years to impose national policy frameworks on the States and intervene in areas of exclusive State jurisdiction. In Australia, where vertical fiscal imbalance (VFI) is considerable, the Commonwealth's Specific Purpose Payments (SPPs) contribute a substantial share of revenue and affect a wide range of policy areas at the State level (Fenna 2008; Banks, Fenna and McDonald, this volume). Both VFI and conditionality are low in Canada, but in the United States 'categorical grants-in-aid' similarly impose a wide range of conditions known as 'sanctions' or 'mandates' (Boadway 2007; Fox 2007). To provide some respite from the centralising effect of that conditionality in various programs, Congress allows the federal administration to grant States 'waivers' authorising them to deviate in permitted ways from the template. In practice as well as principle this allows for some degree of laboratory federalism to be promoted (Weissert and Weissert 2008).

Outliers

Were we to include other, less-conventionally structured, federations such as Spain or Belgium, the diversity would increase substantially. Here it is the EU that is the outlier — a quasi-federation or modern variant of a confederation that has a far lower degree of centralisation than even the most decentralised of the established federation (Bomberg, Peterson and Stubbs 2012; Dinan 2010; Laursen 2011; Majone 2006). While federalism is the most useful lens through which to view the EU (Hueglin and Fenna 2006), there are important caveats. Three features particularly distinguish the EU. First, the EU is made up of sovereign nation-states, most with their own separate languages and national identities. Second, the EU was formed after its members had established their own distinct (and in many cases quite comprehensive) welfare states and the governing assumption was that market integration could proceed without any equivalent social policy integration. And third, the EU has no direct taxing power with which to establish a fiscal hegemony

that could be used to direct policy in areas outside its formal jurisdiction or ‘competence’. Paradoxically, perhaps, these strongly decentralised characteristics guarantee the EU the kind of ‘federal’ character that is steadily being eroded by centralisation in some of the classic federations.

The complexity of modern federal practice

The final point concerns the character of modern federalism — and, in particular, the wide gap between federalism in theory and federalism as it actually exists. Federations have evolved into highly complex and often messy arrangements of political and administrative entanglement that conform only very approximately to ideal-typical models. This is particularly the case for the Anglo federations, where a constitutional division of powers designed in the eighteenth or nineteenth centuries has had to adapt to modern conditions (Fenna 2007). The consequence is a wide range of policy domains where traditional local responsibility has been subject to central government involvement, direction or influence. Taking Australia as a particularly pronounced example, we find the Commonwealth exercising extensive influence in areas that are constitutionally the domain of the States. In an arrangement that is sometimes called ‘cooperative federalism’, the States typically retain administrative responsibility for service delivery but are subject to some form of Commonwealth steering (Williams and MacIntyre 2006). This is accomplished using various mechanisms, but predominant among those are conditional grants (Fenna 2008; Morris 2007) made possible by high levels of vertical fiscal imbalance.

In most federations, social welfare, education and health care have traditionally been a local responsibility but over time those have become ‘nationalised’ to one degree or another and in one form or another. This has happened for a variety of mutually-reinforcing reasons, among them the fact that many now have, or are perceived as having, national dimensions that were absent previously. Traditionally regarded as a matter of almost entirely local import, education has in recent years, for instance, come to be seen as integral to the economic vitality of the nation because of the perceived importance of ‘human capital’ to productivity and innovation.

1.3 What might benchmarking do for federalism?

Given these tendencies and variations in modern federal practice, it is not surprising to find that federalism has affinities with both types of benchmarking. On the one side, federalism and external benchmarking are both about utilising multiple

experiences to identify better ways of doing things; and formalised benchmarking may offer ways to harness the policy learning potential of federalism. On the other side, the increasingly ‘top-down’ nature of some federations creates a natural fit with the focus of internal benchmarking on performance management of subordinate units.

Governments as learning enterprises

If we take the original ‘external’ private sector model of benchmarking where independent firms initiate comparative assessment of their performance as a learning exercise that allows them to incorporate elements of ‘best practice’, then these affinities between benchmarking and federal systems are immediately apparent. We might imagine a federation where the constituent units act like improvement-seeking enterprises, perpetually gauging their performance against fellow governments and incorporating lessons of experience. In this ideal world of policy experimentation and learning, federalism is a ‘laboratory’ for policy improvement and everyone is leveraging themselves up, never reinventing the wheel.

For a variety of reasons the real world is not quite like that.

- Governments have a suboptimal propensity for experimentation.
- Gauging performance and identifying ‘best practice’ is not always easy.
- Policy objectives are often value-laden.
- Governments are under electoral pressure not to engage in open self-assessment.
- Mechanisms for cross-jurisdictional learning may be inadequate.
- In many domains there is an homogenising central government influence.

Seen in this light, the introduction of benchmarking practices and requirements could supply the necessary stimulus and mechanism for competitive improvement and policy learning. Governments that voluntarily enter into benchmarking agreements — ‘benchmarking clubs’ — can create a framework in which more systematic evaluation, greater experimentation, and enhanced learning occurs. Given political realities, this is likely to focus on aspects of service delivery design rather than policy frameworks.

Limitations on the likelihood of sub-national governments engaging in benchmarking of their own volition suggest at least two possible alternatives. One is that benchmarking is done by an independent, non-governmental, institution. This has the advantages that independence brings in the potential for neutral assessment.

It has the advantages being non-governmental brings in not having any impact on the power dynamics of the federal system. However, it also has weaknesses that are the reverse of the advantages: such agencies are at the mercy of governments from whom they seek information; they have no formal leverage. The other alternative, then, is for the task to be executed by an agent with real authority — the central government.

What role for the central government?

In the classic federal model of the Anglo federations, there is little legitimate role for the central government in overseeing the activities of the constituent units — including via benchmarking. Sub-national governments are accountable for their performance to their own voters, not to the national government or the national community. However, reality is far more complex and there are a number of reasons to relax that stricture.

Most importantly, there is the animating and facilitating role that the central government can play in generating an optimal degree of benchmarking between the constituent units in areas where they continue to dominate. Given the obstacles to spontaneous or bottom-up experimentation and learning, there is a constructive role for the central government in encouraging experimentation, promoting and coordinating comparative performance measurement, and facilitating learning. Dorf and Sabel (1998) somewhat grandiosely call this a ‘constitution of democratic experimentalism’ and argue that while the central government must avoid suppressing policy initiative at the subnational level, benign neglect is insufficient. Also in the US context, Metzenbaum (2008) argues that federal agencies overseeing programs delivered by States should adopt a ‘learning and leadership role’ facilitated by performance management using ‘goals, measurement, and incentives’.

There are also the various programs directed and funded to one degree or another by the central government that operate in areas of sub-national jurisdiction. Benchmarking in those contexts represents an alternative mode of coordination that potentially exchanges ‘micro-management’ type controls for a set of incentives that focus on what policy is ultimately all about: outcomes. Potentially, a switch from input and outputs to an outcomes focus would encourage experimentation and learning in the effort to find more effective and efficient means to ends at the service delivery level.

1.4 What do we find?

In practice, we find a wide range of intergovernmental benchmarking experiences whether in unitary countries, between independent countries, or within federal systems — with great variation in both degree and type. At one extreme, the UK provides illustration of the kind of coercive, top-down, performance management that can be executed in a unitary state. At the other extreme, the Organisation for Economic Co-operation and Development (OECD) illustrates the entirely non-coercive benchmarking of sovereign states. Between those two extremes lie federal systems, but even there, practice tends to range along a continuum between the more centralised and the more decentralised cases. In decentralised federations such as Switzerland and Canada we find only modest initiatives and very little of a top-down nature. Insofar as examples are emerging, such as in the Canadian health system (see Baranek, Veillard and Wright, this volume) it may be in areas where a commitment to diversity has diminished (Graefe and Bourns 2009). Three general approaches can be identified: monitoring by independent agencies; top-down, performance monitoring and management; and collegial benchmarking.

Independent monitoring

In a number of countries, performance monitoring of constituent units has been, or is being, done by non-governmental organisations or institutions. In the United States, the Pew Center (2008), carries out a periodic ‘Grading the States’ exercise. Summary assessment is presented in ‘report card’ or ‘league table’ style using a twelve point scale with information made publicly available through website presentation. In their performance assessment of the States, the Pew Center judged the well-integrated use of performance measurement by State governments as, in turn, an important contributor to success (Barrett and Greene 2008; Moynihan and Ingraham 2003). In Germany, a similar assessment has been carried out by the private-sector Bertelsmann Foundation, focusing particularly on fiscal performance (Berthold, Koegel and Kullas 2009; Berthold, Kullas and Müller 2007; Wagschal, Wintermann and Petersen 2009). More recently, the Foundation has expanded its remit to compare governmental performance across the OECD (Bertelsmann Stiftung 2011). In Switzerland, a university-based institute, the Databank on Swiss Cantons and Municipalities, has carried out performance comparisons on a range of fiscal and governance indicators with results publicised via website (Bochsler *et al.* 2004; Koller, Heuberger and Rolland 2011).⁴

⁴ *Base de données des cantons et des villes suisses* (BADAC), at the Institute of Advanced Studies in Public Administration (IDHEAP) in Lausanne.

Such independent monitoring has evident advantages and disadvantages. By their unintrusive nature and apparently disinterested focus on strengths and weaknesses across jurisdictions, such exercises are entirely consonant with federalism and contribute a degree of comparative performance assessment that would otherwise be lacking. This should contribute to both laboratory and competitive federalism. At the same time, though, such non-governmental organisations may well have their own ideological agenda. And, quite separately, there is the question of how much impact they are likely to have. Operating to a large extent with freely available data, independent monitoring may end up measuring things not because they are important or revealing but simply because the data exist and are available. Having no ownership of the exercise, governments may also disregard the findings. The impact of these assessments is unclear.

Top-down Performance Monitoring and Management

At the other extreme from independent monitoring is top-down monitoring where the central government uses internal benchmarking much as a large business enterprise would with its operating units: as a means of driving performance improvement. Such exercises typically represent the continuance in new form of traditional centralising trends in federal systems whereby the national government uses particular constitutional or fiscal levers to achieve a *de facto* alteration in the division of powers and responsibilities. The US case canvassed here (Wong, in this volume) is coercive in nature; recent Australian examples discussed below are much less so.

Benchmarking Swiss employment services

Top-down benchmarking is not to be expected in as decentralised a federation as Switzerland. One failed attempt to impose from the centre, though, illustrates some of the tensions. The introduction of a national scheme of performance management for Switzerland's public employment service followed the logic of combining devolution of managerial responsibility with performance monitoring and sanctioning. Under the Swiss Constitution (Art. 110c), overall responsibility for economic management and specifically for employment services is in the hands of the national government — a power exercised by the Secretary for Economics (SECO). Meanwhile, the actual administration of the relevant services — in this case, employment services — is a cantonal responsibility. Beginning in 2000, SECO installed a system whereby individual performance contracts were signed with each canton; indicators were established; and budgetary rewards scheduled for higher performers (Hilbert 2007).

The program hinged on its system of financial rewards, but such an approach proved difficult and contentious for a number of reasons and was almost immediately abandoned. The indicators measured success but did nothing to guide improvement; the indicators failed to capture any success jurisdictions might have had in *preventing* unemployment in the first place; and publicising adverse findings would cause reputational damage to offices as a consequence of which their ability to engage successfully with employers and the unemployed and thus to ‘perform’ would be further reduced. On top of this was the problem that underperformers were punished by being denied the extra funding that they may well have needed to improve their performance.

No Child Left Behind

The most prominent example of large-scale top-down or internal benchmarking is the US government’s *No Child Left Behind Act* of 2001 (NCLB),⁵ which effected a major shift of control over primary and secondary schooling away from the States who traditionally exercised almost all responsibility in the field (Manna 2006; McGuinn 2006, Wong, this volume). This brought the States explicitly into the performance management fold that Congress had established with passage of the *Government and Performance Results Act* of 1993 requiring federal government agencies to practise performance management. NCLB was unilaterally developed and imposed on the States as an extension of Congress’s traditional conditional (‘categorical’) grant approach to extending its reach to matters within State jurisdiction. Resistance was significant and for reasons pertaining to federalism and to the difficulties of governing by performance measurement, many commentators regard achievements as small (Manna 2011; Radin 2006; Ravitch 2009; Shelly 2008 — though cf. Wong, this volume).

Benchmarking the Australian States

As described below, Australia has long practised a form of collegial benchmarking via the multijurisdictional *Report on Government Services* but that was intensified with the sweeping changes made to Australian federalism in 2008–09 (Fenna 2012; Banks, Fenna and McDonald, this volume). These built on, and worked through, the peak body of Australian federalism, the Council of Australian Governments (COAG). The very large number of existing conditional grant programs were consolidated into a handful of block grants (Treasury 2009) and in exchange for the removal of sundry input conditions, the COAG Reform Council was mandated to

⁵ An Act to Close the Achievement Gap with Accountability, Flexibility, and Choice, so that No Child is Left Behind.

publish performance assessments and carry out benchmarking of State service delivery (see Rimmer, this volume, O’Loughlin, this volume).⁶ Adoption of this performance model represented a concession to a long string of reviews and analyses criticising the way tied grants were being used by the Commonwealth and had been advocated by the States (Allen Consulting Group 2006; Garnaut and FitzGerald 2002; JCPA 1995). It is about ‘letting the managers manage’ (OECD 1997, p. 10), with the ‘managers’ in this case being the State governments and their various agencies. Under the scheme, performance agreements are developed collaboratively and no sanctions are attached. This does not mean that the use of old-style tied grants to intervene in areas of State jurisdiction has been abandoned — indeed, there is some concern about how actively the Commonwealth continues to use that instrument (O’Meara and Faithful 2012). However, it does mean that a substantially more cooperative and outcomes-focused approach is being established in a number of major policy areas. Earlier attempts to introduce performance monitoring in major tied grant programs had run aground on problems of data quality and interpretation (Monro 2003), and it remains to be seen whether this new and more comprehensive attempt will surmount those obstacles.

Collegial benchmarking

The reality is that the governments who are the subject of the benchmarking also need to be the authors in some way of that benchmarking. This is the case in collegial-style monitoring carried out on the basis of intergovernmental agreement and cooperation between jurisdictions whether in unitary or federal systems. Central governments are typically involved, but in a facilitative capacity. The audience may be primarily the governments themselves, or it may be the community more broadly.

The OECD

The most developed example of this is not between jurisdictions within one country, but between independent countries. For more than half a century now, the OECD has sought to promote performance improvement and the adoption of best practice models by benchmarking the performance of its member states (Cotis 2005; Sullivan 1997). Some of that international benchmarking has had noticeable knock-on effect within those member states. Healthcare is one area where this has been the case (OECD 1985; see Baranek, Veillard and Wright, this volume); however, it is in education that the most evident impact has occurred. Through its Programme for

⁶ See the 2009 *Intergovernmental Agreement on Federal Financial Relations* between the Commonwealth and the States and Territories and the *Federal Financial Relations Act 2009*.

International Student Assessment (PISA), the OECD has galvanised education policy across the OECD over the past decade.

More significant in other areas of its work has been the OECD's use of 'peer review' to share its findings and reinforce its message.

Peer review can be described as the systematic examination and assessment of the performance of a State by other States, with the ultimate goal of helping the reviewed State improve its policy making, adopt best practices, and comply with established standards and principles. (Pagani 2002, p. 4)

The operative phrase here is 'comply with' — given that the OECD has no direct leverage whatsoever over the actions of its member states, it 'plays the ideas game' and peer review is one mechanism through which it hopes to win that game. For the OECD, peer review is a 'sort of "soft enforcement" system' (Pagani 2002, p. 12). In general, though, conclusions seem to be that the OECD's impact has generally been very modest — 'the efficacy of OECD recommendations is low' (Armingeon 2004, p. 228; Lodge 2005).

Australia's Report on Government Services

The leading example of collegial benchmarking in federal systems is probably Australia's *Report on Government Services* (RoGS), now in its fifteenth year of publication (see Banks and McDonald, this volume). A steering committee representing all governments establishes the performance monitoring framework and overseas publication of the *Report*; an arm's length research agency of the Commonwealth government, the Productivity Commission, acts as the node of the exercise: serving as secretariat, compiling the data and producing the reports. While RoGS does not cover everything State governments are involved in doing, it does cover an ambitiously wide range of public services making up a substantial part of State government activity. In a number of ways RoGS stands out as exemplary practice; its impact, though, seems to have been modest. This may reflect a broader problem that the wider audience for performance data is not really paying attention (Pollitt 2006).

Sectoral examples

There is no real equivalent to RoGS in other federations, though one can find similar arrangements operating on a sector-specific basis. One notable example in Switzerland is the way that the Confederation facilitates cantonal performance monitoring in the area of sustainability policy through the Office for Spatial Development (ARE). As Wachter (this volume) notes, the main purpose of the

central government's role in this instance is to promote data quality and thus utility — in particular to promote the comparability of data generated on a local basis. Another example is found in Canada, where the federal government acts as a node for a similar exercise in the area of health and hospital services, through CIHI, the Canadian Institute for Health Information (see Baranek, Veillard and Wright, this volume).

From information to learning?

Even though central governments play a role in these collegial benchmarking exercises, that role is a restrained one — generally limited to some combination of instigation and facilitation. On that basis, we can say that collegial benchmarking is close to the private sector model of external benchmarking and thus aligned with the principles of federalism in a way that coercive, top-down, approaches are not. The question to be asked of these various instances of collegial benchmarking is how effectively their increasingly sophisticated generating and aggregating of performance data feeds back into policy learning and service delivery improvement in the individual jurisdictions. In other words, to what extent does performance monitoring actually translate into true benchmarking? The focus in the RoGs, ARE and CIHI cases is on quantitative indicators and comparative performance measurement; mechanisms for qualitative learning are absent or very much secondary. This brings us to the European Union's Open Method of Coordination.

The EU's Open Method of Coordination

The EU has developed the Open Method of Coordination (OMC) as a mode of policy coordination for application in areas where it lacks jurisdiction (Tholoniati 2010). In some ways the OMC is similar to the work of the OECD — reflecting the degree to which the EU lies somewhere between a federation and an international organisation (Casey 2004; Groenendijk 2011; Kröger 2009; Lodge 2005; Schäfer 2006). The OMC is described as a form of 'soft law' in contradistinction to the 'hard law' that the EU exercises via its 'directives'. Prominent among those policy domains where the EU lacks authority are the institutions of the welfare state, education systems and labour market programs — in other words social policy broadly defined. One of the most prominent areas of application has been to 'social inclusion' (Marlier *et al.* 2007; see Vanhercke and Lelie, this volume). Social policy was originally seen as incidental to the EU's main objective of promoting economic dynamism through economic integration but is now regarded as representing essential factors in fiscal and economic performance.

Under the ‘Lisbon Strategy’ proclaimed in 2000, the EU has pursued improvement in those policy areas by establishing performance measurement and benchmarking frameworks; engaging in evaluation and peer review; and encouraging mutual learning. With a peer review system rather less prescriptive than the OECD’s, the OMC was designed to have strong and complementary quantitative and qualitative dimensions, to be voluntary, and to promote ‘contextualised learning’ — that is, learning based on recognition of the different circumstances and different cultural and institutional orders prevailing in different jurisdictions. This has been hailed as representing a breakthrough in experimentalist governance (Sabel and Zeitlin 2010).

Dissatisfaction with the OMC’s limited impact led after a few years to the recommendation that it switch to a ‘naming, shaming and faming’ approach in the form of league tables that would more aggressively cajole Member States into adopting best practices (Kok 2004). That recommendation was rejected and there is little reason to think that it would have been successful. While the OMC epitomises the federal principle of cooperation, mutual respect, autonomous accountability and improvement through mutual learning, its substantive impact remains much debated (Heidenreich and Zeitlin 2009; Kerber and Eckardt 2007; Kröger 2009; Radaelli 2008). In particular, there is the widespread view that among other things its lack of teeth and the embedded differences in the policy regimes across the EU, render it ineffectual. However, this may reflect unrealistic expectations and be insensitive to more subtle and incremental ways in which the OMC works (see Vanhercke and Lelie, this volume).

1.5 Conclusion

Federalism and benchmarking are enjoying a tentative, exploratory, relationship that is partly based in good faith attempts to fulfil some of federalism’s potential as a learning-oriented governance arrangement and partly reflective of long-running centralisation dynamics. Derived from the private sector, benchmarking has been championed as a way to infuse public sector organisations with a stronger focus on both efficiency and results. Both the private sector’s voluntary ‘external’ benchmarking and its mandatory ‘internal’ benchmarking have their public sector equivalents. The wide variety in federal systems means that variants of both external and internal types can be found, ranging from the more top-down and coercive internal types to the ‘bottom up’ external types functioning on a collegial basis and oriented more to learning. The latter are more compatible with the federal idea while the former reflect the realities of some contemporary federal systems.

Lacking the same incentives as business firms, and facing a number of disincentives particular to the public sector, governments are cautious about participating in

comparative performance measurement and analysis. It is not surprising, then, that some of the main examples correspond more closely to the internal, top-down, model. Where those have been imposed unilaterally and carry sanctions they are likely to exacerbate dysfunctional elements of both federalism and benchmarking — in no small part because effective benchmarking relies on reliable feedback processes. Where such arrangements have been developed collaboratively and rely minimally on sanctions, benchmarking may offer an administratively and substantively superior alternative to more directive modes of centralised policy making in federal systems.

Examples of external, ‘collegial’, benchmarking in federal systems are limited. Successful examples rely on iterative development and confidence-building. There is almost always an important role for central governments in instigating and facilitating such exercises — providing incentives to participate; acting as an information node promoting comparability, collection and synthesis of data. These tend to be found in more decentralised federations, or indeed the EU, which is so decentralised as to be not yet a federation in the conventional sense and where the great diversity of membership places a premium on ‘contextualised comparison’.

References

- AC (Audit Commission) 2009, *Nothing But the Truth?*, 5 November 2009, http://www.audit-commission.gov.uk/localgov/nationalstudies/Pages/nothingbutthetruth_copy.aspx%3E.
- Allen Consulting Group 2006, *Governments Working Together? Assessing Specific Purpose Payment arrangements*, Government of Victoria, Melbourne.
- Allen, W.H. 1907, *Efficient Democracy*, Dodd, Mead and Company, New York.
- Armingeon, K. 2000, ‘Swiss Federalism in Comparative Perspective’, in U. Wachendorfer-Schmidt (ed.), *Federalism and Political Performance*, Routledge, London.
- 2004, ‘OECD and National Welfare State Development’, in K. Armingeon and M. Beyeler (eds), *The OECD and European Welfare States*, Edward Elgar, Cheltenham, pp. 226–41.
- Atkinson, T., Cantillon, B., Marlier, E. and Nolan, B. 2002, *Social Indicators: the EU and social inclusion*, Oxford University Press, Oxford.
- Barrett, K. and Greene, R. 2008, ‘Measuring Performance: the State management report card for 2008’, *Governing*, March.
- Bauer, R.A. (ed.) 1966, *Social Indicators*, MIT Press, Cambridge MA.

-
- Bertelsmann Stiftung (ed.) 2011, *Sustainable Governance Indicators 2011: policy performance and governance capacities in the OECD*, Bertelsmann Stiftung, Gütersloh.
- Berthold, N., Koegel, D. and Kullas, M. 2009, *Die Bundesländer im Innovationswettbewerb*, Bertelsmann Stiftung, Gütersloh.
- Berthold, N., Kullas, M. and Müller, A. 2007, *Die Bundesländer im Standortwettbewerb*, Bertelsmann Stiftung, Gütersloh.
- Bevan, G. and Hood, C. 2006, 'What's Measured is What Matters: targets and gaming in the English public health care system', *Public Administration*, vol. 84, no. 3, pp. 517–38.
- Boadway, R. 2007, 'Canada', in A. Shah (ed.), *The Practice of Fiscal Federalism: comparative perspectives*, McGill–Queen's University Press, Montreal and Kingston.
- Bochsler, D., Koller, C., Sciarini, P., Traimond, S. and Trippolini, I. 2004, *Les cantons suisse sous la loupe: autorités, employés publics, finances*, Haupt, Berne.
- Bomberg, E., Peterson, J. and Stubbs, A. 2012, *The European Union: how does it work?*, 3rd edn, Oxford University Press, Oxford.
- Braun, D. 2010, 'Multi-level Governance in Germany and Switzerland', in H. Enderlein, S. Wälti and M. Zürn (eds), *Handbook on Multi-level Governance*, Edward Elgar, Cheltenham.
- Bryce, J. 1893, *The American Commonwealth*, 3rd edn, vol. 1, 2 vols., Macmillan, London.
- Bunge, M. 1975, 'What is a Quality of Life Indicator?', *Social Indicators Research*, vol. 2, no. 1, pp. 65–79.
- Burgess, S., Wilson, D. and Worth, J. 2010, *A Natural Experiment in School Accountability: the impact of school performance information on pupil progress and sorting*, Centre for Market and Public Organisation, Bristol Institute of Public Affairs, Bristol.
- Carter, N., Klein, R. and Day, P. 1995, *How Organisations Measure Success: the use of performance indicators in government*, Routledge, London.
- Casey, B.H. 2004, 'The OECD Jobs Strategy and the European Employment Strategy: two views of the labour market and the welfare state', *European Journal of Industrial Relations*, vol. 10, no. 3, pp. 329–52.
- Cotis, J.P. 2005, 'OECD'S Role in Global Governance: the contribution of structural indicators and benchmarking', in OECD (ed.), *Statistics, Knowledge*

and Policy: key indicators to inform decision making, Organisation for Economic Co-operation and Development, Paris.

DCLG (Department for Communities and Local Government) 2009, *National Indicators for Local Authorities and Local Authority Partnerships: updated national indicator definitions*, Department for Communities and Local Government, London.

— 2010, ‘Eric Pickles to disband Audit Commission in new era of town hall transparency’, Department for Communities and Local Government, London, 13 August.

Department of Health Education and Welfare 1970, *Toward a Social Report*, University of Michigan Press, Ann Arbor MI.

Dinan, D. 2010, *Ever Closer Union: an introduction to European integration*, 4th edn, Lynne Rienner, Boulder CO.

Dorf, M.C. and Sable, C.F. 1998, ‘A Constitution of Democratic Experimentalism’, *Columbia Law Review*, vol. 98, no. 2, pp. 267–473.

Downe, J. 2008, ‘Inspection of Local Government Services’, in H. Davis and S. Martin (eds), *Public Services Inspection in the UK*, Jessica Kingsley, London, pp. 19–36.

Esping-Andersen, G. 2005, ‘Indicators and Social Accounting for 21st Century Social Policy’, in OECD (ed.), *Statistics, Knowledge and Policy: key indicators to inform decision making*, Organisation for Economic Co-operation and Development, Paris.

Fenna, A. 2007, ‘The Malaise of Federalism: comparative reflections on Commonwealth–State Relations’, *Australian Journal of Public Administration*, vol. 66, no. 3, pp. 298–306.

— 2008, ‘Commonwealth Fiscal Power and Australia Federalism’, *University of New South Wales Law Journal*, vol. 31, no. 2, pp. 509–29.

— 2012, ‘Adaptation and Reform in Australian Federalism’, in P. Kildea, A. Lynch and G. Williams (eds), *Tomorrow’s Federation: reforming Australian government*, Federation Press, Leichhardt NSW, pp. 26–42.

Fox, W. 2007, ‘United States of America’, in A. Shah (ed.), *The Practice of Fiscal Federalism: comparative perspectives*, McGill–Queen’s University Press, Montreal and Kingston.

Garnaut, R. and FitzGerald, V. 2002, *Review of Commonwealth–State Funding: final report*, Committee for the Review of Commonwealth–State Funding, Melbourne.

-
- Goddard, M. and Mannion, R. 2004, 'The Role of Horizontal and Vertical Approaches to Performance Measurement and Improvement in the UK Public Sector', *Public Performance and Management Review*, vol. 28, no. 1, pp. 75–95.
- Goldstein, H. and Leckie, G. 2008, 'School League Tables: what can they tell us?', *Significance*, vol. 5, no. 2, pp. 67–9.
- Graefe, P. and Bourns, A. 2009, 'The Gradual Defederalization of Canadian Health Policy', *Publius*, vol. 39, no. 1, pp. 187–209.
- Grizzle, G. A. 2002, 'Performance Measurement and Dysfunction', *Public Performance and Management Review*, vol. 25, no. 4, pp. 363–69.
- Groenendijk, N.S. 2011, 'EU and OECD Benchmarking and Peer Review Compared', in F. Laursen (ed.), *The EU and Federalism: politics and policies compared*, Ashgate, Farnham, pp. 181–202.
- Hatry, H.P. 2006, *Performance Measurement: getting results*, 2nd edn, Urban Institute, Washington DC.
- Heidenreich, M. and Zeitlin, J. (eds) 2009, *Changing European Employment and Welfare Regimes: the influence of the open method of coordination on national reforms*, Routledge, London.
- Hilbert, C. 2007, 'Implementation of Performance Measurement in Public Employment Services in Switzerland', in J. De Koning (ed.), *The Evaluation of Active Labour Market Policies: measures, public private partnerships and benchmarking*, Edward Elgar, Cheltenham.
- Hood, C. 2006, 'Gaming in Target World: the targets approach to managing British public services', *Public Administration Review*, vol. 66, no. 4, pp. 515–22.
- Hueglin, T.O. and Fenna, A. 2006, *Comparative Federalism: a systematic inquiry*, Broadview Press, Peterborough ON.
- Hvinden, B. 2005, 'Social Inclusion – Do We Have the Right Indicators?', in OECD (ed.), *Statistics, Knowledge and Policy: key indicators to inform decision making*, Organisation for Economic Co-operation and Development, Paris.
- Innes, J.E. 1990, *Knowledge and Public Policy: the search for meaningful indicators*, 2nd edn, Transaction Publishers, New Brunswick NJ.
- JCPA (Joint Committee of Public Accounts) 1995, *The Administration of Specific Purpose Payments: a focus on outcomes*, Parliament of the Commonwealth of Australia, Canberra.
- Johnsen, Å. 2008, 'Performance Information and Educational Policy Making', in W. Van Dooren and S. Van de Walle (eds), *Performance Information in the Public Sector: how it is used*, Palgrave Macmillan, Basingstoke, pp. 157–73.

-
- Kelman, S. and Friedman, J.N. 2009, 'Performance Improvement and Performance Dysfunction: an empirical examination of impacts of the emergency room wait-time target in the English National Health Service', *Journal of Public Administration Research and Theory*, vol. 19, no. 4, pp. 917–46.
- Kerber, W. and Eckardt, M. 2007, 'Policy Learning in Europe: the open method of co-ordination and laboratory federalism', *Journal of European Public Policy*, vol. 14, no. 2, pp. 227–47.
- Kok, W. 2004, *Facing the Challenge: the Lisbon Strategy for Growth and Employment*, European Communities, Brussels, <europa.eu.in/comm/lisbon_strategy/index_en.html>.
- Koller, C., Heuberger, N. and Rolland, A.C. 2011, *Staatsmonitoring 1990–2011: indikatoren zur messung der öffentlichen verwaltung und der behörden auf kantonaler ebene*, IDHEAP, Lausanne.
- Kramer, J. 2005, 'Federal Republic of Germany', in J. Kincaid and G.A. Tarr (eds), *Constitutional Origins, Structure, and Change in Federal Countries*, McGill–Queen's University Press, Montreal and Kingston.
- Kravchuk, R.S. and Schack, R.W. 1996, 'Designing Effective Performance Measurement Systems under the Government Performance and Results Act', *Public Administration Review*, vol. 56, no. 4, pp. 348–58.
- Kröger, S. 2009, 'The Effectiveness of Soft Governance in the Field of European Anti-Poverty Policy: operationalization and empirical evidence', *Journal of Comparative Policy Analysis*, vol. 11, no. 2, pp. 197–211.
- Laursen, F. (ed.) 2011, *The EU and Federalism: politics and policies compared*, Ashgate, Farnham.
- Linder, W. 2010, *Swiss Democracy: possible solutions to conflict in multicultural societies*, 3rd edn, Palgrave Macmillan, Basingstoke.
- Lodge, M. 2005, 'The Importance of Being Modern: international benchmarking and national regulatory innovation', *Journal of European Public Policy*, vol. 12, no. 4, pp. 649–67.
- Majone, G. 2006, 'Federation, Confederation, and Mixed Government: a EU–US comparison', in A. Menon and M. Schain (eds), *Comparative Federalism: the European Union and the United States in comparative perspective*, Oxford University Press, New York, pp. 121–48.
- Manna, P. 2006, *School's In: federalism and the national education agenda*, Georgetown University Press, Washington, D.C.
- 2011, *Collision Course: federal education policy meets State and Local realities*, CQ Press, Washington DC.

-
- Marlier, E., Atkinson, A.B., Cantillon, B. and Nolan, B. 2007, *The EU and Social Inclusion: facing the challenges*, Policy Press, Bristol.
- McGuinn, P.J. 2006, *No Child Left Behind and the Transformation of Federal Educational Policy, 1965–2005*, University Press of Kansas, Lawrence KS.
- McLean, I., Haubrich, D. and Gutiérrez-Romero, R. 2007, ‘The Perils and Pitfalls of Performance Measurement: the CPA regime for Local Authorities in England’, *Public Money and Management*, vol. 27, no. 2, pp. 111–8.
- Metzenbaum, S.H. 2008, ‘From Oversight to Insight: Federal agencies as learning leaders in the information age’, in T.J. Conlan and P.L. Posner (eds), *Intergovernmental Management for the Twenty-First Century*, Brookings Institution Press, Washington DC, pp. 209–42.
- Monro, D. 2003, ‘The Role of Performance Measures in a Federal–state Context: the examples of housing and disability services’, *Australian Journal of Public Administration*, vol. 62, no. 1, pp. 70–9.
- Morris, A. 2007, ‘Commonwealth of Australia’, in A. Shah (ed.), *The Practice of Fiscal Federalism: comparative perspectives*, McGill–Queen’s University Press, Montreal and Kingston.
- Moynihan, D.P. and Ingraham, P.W. 2003, ‘Look for the Silver Lining: when performance-based accountability systems work’, *Journal of Public Administration Research and Theory*, vol. 13, no. 4, pp. 469–90.
- Nutley, S. and Smith, P.C. 1998, ‘League Tables for Performance Improvement in Health Care’, *Journal of Health Services Research and Policy*, vol. 3, no. 1, pp. 50–7.
- O’Meara, P. and Faithful, A. 2012, ‘Increasing Accountability at the Heart of the Federation’, in P. Kildea, A. Lynch and G. Williams (eds), *Tomorrow’s Federation: reforming Australian government*, Federation Press, Leichhardt NSW, pp. 92–112.
- OECD 1982, *Social Indicators*, Organisation for Economic Co-operation and Development, Paris.
- 1985, *Measuring Health Care, 1960-1983: expenditure, costs and performance*, Organisation for Economic Co-operation and Development, Paris.
- 1997, *In Search of Results: performance management practices*, Organisation for Economic Co-operation and Development, Paris.
- Oeter, S. 2006, ‘Federal Republic of Germany’, in K. Le Roy and C. Saunders (eds), *Legislative, Executive, and Judicial Governance in Federal Countries*, McGill–Queen’s University Press, Montreal and Kingston.

-
- Osborne, D. and Gaebler, T. 1992, *Reinventing Government: how the entrepreneurial spirit is transforming the public sector*, Addison-Wesley, Reading MA.
- Pagani, F. 2002, *Peer Review: a tool for co-operation and change — an analysis of an OECD working method*, Organisation for Economic Co-operation and Development, Paris.
- Pew Center 2008, *Grading the States 2008*, The Pew Center on the States, Washington DC, <http://www.pewcenteronthestates.org/gpp_report_card.aspx%3E.
- Pollitt, C. 1990, *Managerialism in the Public Services: the Anglo-American experience*, Blackwell, Oxford.
- 2006, ‘Performance Information for Democracy: the missing link’, *Evaluation*, vol. 12, no. 1, pp. 39–56.
- Radaelli, C.M. 2008, ‘Europeanization, Policy Learning, and New Modes of Governance’, *Journal of Comparative Policy Analysis*, vol. 10, no. 3, pp. 239–54.
- Radin, B.A. 2006, *Challenging the Performance Movement: accountability, complexity, and democratic values*, Georgetown University Press, Washington DC.
- Radnor, Z. 2008, ‘Hitting the Target and Missing the Point? Developing an understanding of organizational gaming’, in W. Van Dooren and S. Van de Walle (eds), *Performance Information in the Public Sector: how it is used*, Palgrave Macmillan, Basingstoke, pp. 94–105.
- Ravitch, D. 2009, ‘Time to Kill ‘No Child Left Behind’’, *Education Week*, 8 June.
- Risberg, T.F. 2011, ‘National Standards and Tests: the worst solution to America’s educational problems ... except for all the others’, *George Washington Law Review*, vol. 79, no. 3.
- Sabel, C.F. and Zeitlin, J. (eds) 2010, *Experimentalist Governance in the European Union: towards a new architecture*, Oxford University Press, New York.
- Salmon, P. 2006, ‘Horizontal Competition among Governments’, in E. Ahmad and G. Brosio (eds), *Handbook of Fiscal Federalism*, Edward Elgar, Cheltenham.
- Sanger, M.B. 2008, ‘From Measurement to Management: breaking through the barriers to State and Local performance’, *Public Administration Review*, vol. 68, no. s1, pp. S70–S85.
- Saunders, C. 2005, ‘Commonwealth of Australia’, in J. Kincaid and G.A. Tarr (eds), *Constitutional Origins, Structure, and Change in Federal Countries*, McGill–Queen’s University Press, Montreal and Kingston.

-
- Schäfer, A. 2006, 'A New Form of Governance? Comparing the Open Method of Co-ordination to multilateral surveillance by the IMF and the OECD', *Journal of European Public Policy*, vol. 13, no. 1, pp. 70–88.
- Schmitt, N. 2005, 'Swiss Confederation', in J. Kincaid and G.A. Tarr (eds), *Constitutional Origins, Structure and Change in Federal Countries*, McGill–Queen's University Press, Montreal and Kingston.
- Schneider, H.P. 2006, 'The Federal Republic of Germany', in A. Majeed, R.L. Watts and D. Brown (eds), *Distribution of Powers and Responsibilities in Federal Countries*, McGill–Queen's University Press, Montreal and Kingston.
- Shelly, B. 2008, 'Rebels and Their Causes: State resistance to No Child Left Behind', *Publius*, vol. 38, no. 3, pp. 444–68.
- Smith, P. 1995, 'On the Unintended Consequences of Publishing Performance Data in the Public Sector', *International Journal of Public Administration*, vol. 18, no. 2/3, pp. 277–310.
- Sullivan, S. 1997, *From War to Wealth: 50 years of innovation*, Organisation for Economic Co-operation and Development, Paris.
- Tholoniati, L. 2010, 'The Career of the Open Method of Coordination: lessons from a 'soft' EU instrument', *West European Politics*, vol. 33, no. 1, pp. 93–117.
- Treasury, Department of the 2009, *Budget Paper No. 3: Australia's federal relations 2009–10*, Commonwealth of Australia, Canberra.
- Wagschal, U., Wintermann, O. and Petersen, T. 2009, *Konsolidierungsstrategien der Bundesländer: verantwortung für die zukunft*, Bertelsmann Stiftung, Gütersloh.
- Watson, G.H. 1994, 'A Perspective on Benchmarking', *Benchmarking: an international journal*, vol. 1, no. 1, pp. 5–10.
- Weissert, C.S. and Weissert, W.G. 2008, 'Medicaid Waivers: license to shape the future of fiscal federalism', in TJ Conlan and PL Posner (eds), *Intergovernmental Management for the Twenty-First Century*, Brookings Institution Press, Washington DC, pp. 157–75.
- West, A. 2010, 'High Stakes Testing, Accountability, Incentives and Consequences for English schools', *Policy and Politics*, vol. 38, no. 1, pp. 23–39.
- Wholey, J.S. and Hatry, H.P. 1992, 'The Case for Performance Monitoring', *Public Administration Review*, vol. 52, no. 6, pp. 604–10.
- Williams, J.M. and MacIntyre, C. 2006, 'Commonwealth of Australia', in A. Majeed, R.L. Watts and D.M. Brown (eds), *Distribution of Powers and Responsibilities in Federal Countries*, McGill–Queen's University Press, Montreal and Kingston.