



Australian Government
Productivity Commission



Benchmarking in Federal Systems

Roundtable Proceedings

Melbourne, 19-20 December 2010

Edited by:
Alan Fenna & Felix Knüpling

© Commonwealth of Australia 2012

ISBN 978-1-74037-406-4

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, the work may be reproduced in whole or in part for study or training purposes, subject to the inclusion of an acknowledgment of the source. Reproduction for commercial use or sale requires prior written permission from the Productivity Commission. Requests and inquiries concerning reproduction and rights should be addressed to Media and Publications (see below).

This publication is available from the Productivity Commission website at www.pc.gov.au. If you require part or all of this publication in a different format, please contact Media and Publications.

Publications Inquiries:

Media and Publications
Productivity Commission
Locked Bag 2 Collins Street East
Melbourne VIC 8003

Tel: (03) 9653 2244
Fax: (03) 9653 2303
Email: maps@pc.gov.au

General Inquiries:

Tel: (03) 9653 2100 or (02) 6240 3200

An appropriate citation for this paper is:

Productivity Commission and Forum of Federations 2012, *Benchmarking in Federal Systems*, Roundtable Proceedings, Melbourne, 19–20 December 2010, eds A. Fenna and F. Knüpling, Productivity Commission, Canberra.

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au) or by contacting Media and Publications on (03) 9653 2244 or email: maps@pc.gov.au

Foreword

Federal systems of government generally give rise to greater coordination problems and transaction costs than unitary states. Yet they can also yield important benefits, through the opportunities they provide for policy experimentation and learning across constituent jurisdictions. However, the latent potential for such inter-jurisdictional learning cannot be fully realised without relevant and accessible information. Australia has made some significant advances in this respect, but so too have other countries, which provides an opportunity for further learning at the international level.

The Productivity Commission was therefore pleased to join with the Forum of Federations to hold an international roundtable on *Benchmarking in Federal Systems*. The Roundtable took place in Melbourne in late 2010, bringing together government officials, academics and practitioners from Australia and a number of other countries. This volume provides updated and elaborated versions of the papers presented at the Roundtable.

The Roundtable provided the opportunity to compare Australian approaches to ‘benchmarking’ with those followed in the United States, Canada, the United Kingdom and Germany, as well as at the European Union level. An introductory chapter by Professor Alan Fenna, who also co-edited the volume, provides a conceptual framework against which the different approaches can be considered.

The Commission is grateful to the Forum of Federations for the invitation to co-host the Roundtable, and to the participants, whose contributions made the Roundtable, and the production of this volume, a valuable exercise.

Gary Banks AO
Chairman

June 2012

Contents

Foreword	iii
Introduction	1
<i>Felix Knüpling</i>	
1 Benchmarking in federal systems	11
<i>Alan Fenna</i>	
Part I: International contributions	
2 From the improvement end of the telescope: benchmarking and accountability in UK local government	41
<i>Clive Grace</i>	
3 The implementation of the <i>No Child Left Behind Act</i>: toward performance-based federalism in US education policy	61
<i>Kenneth K. Wong</i>	
4 Benchmarking health care in federal systems: the Canadian experience	89
<i>Patricia Baranek, Jeremy Veillard and John Wright</i>	
5 Towards benchmarking in Germany	111
<i>Gottfried Konzendorf and Rabea Hathaway</i>	
6 Benchmarking sustainable development in the Swiss confederation	123
<i>Daniel Wachter</i>	
7 Benchmarking social Europe a decade on: demystifying the OMC's learning tools	145
<i>Bart Vanhercke and Peter Lelie</i>	
Part II: Australian contributions	
8 Australia's federal context	187
<i>Gary Banks, Alan Fenna and Lawrence McDonald</i>	

9	Benchmarking and Australia’s Report on Government Services	199
	<i>Gary Banks and Lawrence McDonald</i>	
10	Benchmarking, competitive federalism and devolution: how the COAG reform agenda will lead to better services	227
	<i>Ben Rimmer</i>	
11	Benchmarking and accountability: the role of the COAG Reform Council	247
	<i>Mary Ann O’Loughlin</i>	
12	Australian perspectives on benchmarking: the case of school education	267
	<i>Peter Dawkins</i>	
13	Benchmarking in federal systems: the Queensland experience	281
	<i>Sharon Bailey and Ken Smith</i>	
Dinner address		
14	Commonwealth–State benchmarking and the state of Australian Federalism	295
	<i>Helen Silver</i>	
Appendixes		
A	Roundtable program	301
B	Roundtable participants	303

Introduction

Felix Knüpling¹

Forum of Federations

This volume seeks to shed light on the varying experiences of federal or federal-type systems experimenting with benchmarking as a technique to improve performance and foster learning across their constituent units.² Benchmarking arrangements are being widely adopted across federal systems. All federations face the issue of balancing the interests of the central government³ in key areas of public policy with the desire of constituent units to have autonomy or at least flexibility in terms of how they manage major programs. As part of the ‘new public management’ agenda and the drive towards evidence based policies, benchmarking is emerging as a way of escaping some of the rigidities of traditional conditional grant programs or injecting a new dynamism into federal practices, as well as shifting the focus to outcomes achievement and ‘best practice’.

The contributions covered in this volume have been presented at a conference the Forum of Federations held in cooperation with the Productivity Commission of the Australian Government in October 2010 in Melbourne, bringing together experts and government representatives from both orders of government in Australia as well as from five countries and the European Union. The objective was to share experiences and discuss the applicability of benchmarking exercises in a federal context.

As part of a multi-year research and knowledge-exchange program run by the Forum of Federations on ‘Benchmarking in Federal Systems’, this volume aims to fill a knowledge gap. A lot of research has been conducted on the techniques of benchmarking already, but, as of yet, there has been no systematic comparison drawing out comparative experiences or lessons learnt with a special focus on federal systems.⁴

¹ Felix Knüpling is the Head of Programs and Partnerships at the Forum of Federations.

² The *States, Provinces, Länder* or *Cantons* of which the federation is constituted.

³ This book uses both ‘central’ government as well as ‘federal’ to refer to the government with national responsibilities.

⁴ To my knowledge, there are only two major comparative studies which are related to this topic:
1) The OECD has published a report on *Promoting Performance: Using Indicators to Enhance*

The Forum of Federations began its program in 2008, when it was invited to provide international expertise to the German Government, when Germany was going through a process of constitutional reform. Accordingly, this program is designed both for the academic world as well as for those who are involved in shaping or executing policies, law-makers and civil servants. It intends to present and identify comparative experiences that could inform and stimulate ongoing debates on benchmarking in other federal countries.

This volume examines current practices and identifiable trends in Australia, Canada, Germany, the United Kingdom, the United States, Switzerland and the European Union (EU). This list of countries contains five classical federal countries, but also one unitary state (the UK) and a quasi-federal international organisation (the EU). In the case of the UK, the focus is primarily on national–local relationships, while in the case of the European Union it is on relations between the European Commission and the EU member states.

Special attention is given to the Australian case. Australia has been the first federation to systematically employ benchmarking techniques in its intergovernmental relations and can look back to almost two decades of experiences in this area. For this reason the book devotes one section to the Australian experience and one to the ‘international’ case studies.

Every country or political system is, of course, different in nature and any comparisons or generalisations are always difficult to justify. It is important to keep in mind the contexts of the different cases covered in this volume. What works in one case is not automatically applicable in another. However, this should not preclude us from learning from one another — be it from ‘good’ or from ‘not so good’ experiences. It is with this intention that this book has been produced: to open perspectives through an international comparative exercise and to stimulate thinking outside the conventional boxes.

We invited both academic experts as well as government representatives to contribute to the volume. Thematically, the focus of the cases covered in this volume is on the implications of using benchmarking as an alternative to existing modes of coordination in federal systems, as well as on the political and

the Effectiveness of Sub Central Spending (Working Paper 5, Fiscal Relations Network (<http://www.oecd.org/dataoecd/43/56/40832141.pdf>).

2) The Council of Europe’s Committee of Local and Regional Democracy commissioned a comparative study on ‘Performance Management at Local Level’ in its member countries in 2005 (<https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=1163686&SecMode=1&DocId=1345000&Usage=2>).

administrative dimensions of implementation — how to design and operate benchmarking arrangements successfully.

A number of questions arise for federal systems.

- How do benchmarking arrangements affect intergovernmental relations and the functioning of the federal system?
- To what extent might benchmarking practices enhance federalism and what form of benchmarking is most conducive to effective federal practice?
- What are the challenges in moving from performance monitoring to active policy learning?
- Does benchmarking actually lead to improved outcomes?

Authors contributing to this volume were asked to address these questions. They do not give, however, an encompassing overview of federal benchmarking exercises in their country. As experts in specific policy fields they present and discuss individual benchmarking examples and how they relate to and/or are affected by federalism. Given that in many benchmarking cases it is still ‘early days’, it will be no surprise that contributors can often advance only tentative answers to some of these questions.

Contributions in this volume

The book starts off with *Alan Fenna’s* overview on ‘Benchmarking in Federal Systems’. *Fenna* bases his observations not only on the contributions of this volume, but also on the research he was mandated to carry out for the Forum of Federations on this subject.⁵ He provides a conceptual clarification of the main variables discussed in this volume, ‘federalism’ and ‘benchmarking’, and how they relate to each other. In characterising the main differences between benchmarking designs in federal systems along a continuum from top down/coercive benchmarking to bottom up/consensus benchmarking his chapter also lays down the analytical framework for the subsequent chapters.

Part I: International contributions

Following this analytical framework, the cases covered in the next two chapters can be categorised as top-down or coercive benchmarking, with the main objective to increase accountability in specific areas of public policy. In chapter 2, *Clive Grace*

⁵ <http://www.forumfed.org/en/global/thematic/benchmarking.php>.

analyses the changing nature of local government benchmarking in the UK, most visibly noticeable in the abolition of the principal benchmarking and performance management regime for local government in England, the Comprehensive Area Assessment, and the main actor of that regime, the Audit Commission, in 2010–11. Despite some scepticism about the added value of local government benchmarking in the UK, Grace argues that performance assessment will remain relevant in the context of UK public policy. He concludes by drawing out the underlying fundamental relationship that connects benchmarking to service improvement, and suggests a rudimentary framework through which policy makers and practitioners should approach the use of benchmarking methods.

Chapter 3 looks at federal benchmarking in the area of education in the US. *Kenneth Wong* describes the 2001 *No Child Left Behind Act* (NCLB), introduced during the second term of US president George W. Bush to improve the quality of primary education in the United States, as the largest federal benchmarking exercise in the US. The author argues that NCLB, marked the beginning of a serious effort toward performance-based federalism. For some, he writes, NCLB has changed the terms of intergovernmental relations in the US to such an extent as to represent a ‘regime change’. However, as NCLB evolves further under the Obama administration, it remains to be seen whether and if so how the performance-based paradigm will be fully institutionalised in the intergovernmental policy system and to what extent it will continue to produce intergovernmental conflicts, Wong concludes.

Gottfried Konzendorf writes about the emerging process of public policy benchmarking in Germany in chapter 4. Compared to all other cases covered in this volume, Germany is unique in the sense that benchmarking of public service delivery of the *Länder* was made a constitutional provision during a recent constitutional amendment in the context of a major overhaul of German federalism. Thus, this chapter is about the attempt to introduce a holistic federal benchmarking exercise driven through a constitutional provision, and not so much about benchmarking in a specific policy area or sector. Compared to other case covered in this volume benchmarking is still in its infancy in Germany. It is also facing resistance, particularly on the side of the *Länder*. Still, *Konzendorf* expects that further benchmarking exercises in various policy fields will be launched soon and that in the medium term these projects will improve the policy coordination of the two orders of government in Germany.

Taking the decentralised nature of Canadian federalism, and its different legal tradition a constitutionally enshrined provision for inter-provincial benchmarking would be unimaginable in Canada, the focus of chapter 5 by *Patricia Baranek*, *Jeremy Veillard* and *John Wright*. Not surprisingly, the Canadian federal government has no overall strategy to introduce benchmarking in the

intergovernmental context. However, there is a very extensive benchmarking exercise going on in health policy, which the authors attribute to intergovernmental events in the 1990s and the move of policy decision-makers to introduce a performance assessment framework to ensure accountability and efficiency in the health care sector. Although initially Canadian constituent units were rather reluctant to embark on this journey, benchmarking in the health sector is expanding through the comparison of performance with peer groups and the learning from better performers, explain the authors. They conclude by providing an outlook on what requirements are needed to improve the current benchmarking framework further.

Like Canada, Switzerland is also a decentralised federation, as *Daniel Wachter* notes in chapter 6 on benchmarking to promote sustainable development. This is a bottom-up benchmarking regime where a federal agency — the Federal Office for Spatial Development (ARE) — plays a facilitative and cooperative but not a directing role. The collaborative and participative nature of the project is integral to its success, explains *Wachter*. He acknowledges that since participating entities do not have to fear punishment because of inferior performance compared to other the regime has been able to expand in the numbers of participants. The collaborative nature is emblematic of the extent to which this regime is more about learning and sharing best practices in a specific policy area than about exerting some sort of control by the federal government over the use of financial transfers. Thus, it can be considered a ‘soft’ benchmarking regime, since there is no ‘hard’ legislation involved.

The same applies to the European Union’s ‘Open Method of Coordination’ (OMC) on social protection and social inclusion, as *Bart Vanhercke* and *Peter Lelie* write in chapter 7. Although the EU is not a federation, it exhibits important federal features and is therefore relevant for our purposes here (see *Fenna*, this volume). The OMC is designed as communication process of performance assessment of the social policies of EU member states according to common indicators and objectives. In its set up it also falls in the category of what *Fenna* calls collegial benchmarking. The chapter on the OMC provides an overview of a range of benchmarking tools and the way a variety of EU and domestic actors are involved in them. The OMC has been applied by European institutions and stakeholders as a mechanism for coordinating domestic policies in a range of issue areas for which the EU has no formal authority, but also for monitoring and supplementing EU legislative instruments. The authors argue that this benchmarking regime has in effect not been as toothless as many critics have argued. It has evolved as an instrument for learning that proves to be of value for decision makers.

Part II: Australian contributions

The following chapter sets the stage for the contributions on the impact of benchmarking regimes within the Australian federal system. *Gary Banks, Alan Fenna* and *Lawrence McDonald* describe the underlying features of Commonwealth–State relations, with a particular view as to how they relate to benchmarking. They touch on constitutional provisions, Australian fiscal federalism, the cooperative nature of Australian federalism and its high degree of centralisation as well as recent reform steps in the area of fiscal federalism that provide the background for new benchmarking regimes. It is the move away from strict controls over federal transfers to the constituent units that provides the main motive for benchmarking regimes with a focus on outcome performance.

In chapter 9, *Gary Banks* and *Lawrence McDonald* describe and analyse benchmarking in the context of the annual *Report on Government Services* (RoGS) as carried out by the Productivity Commission. Similar to our Swiss case study, RoGS is a collaborative and consensual exercise in which the Commonwealth government plays a facilitative role rather than a directive or coercive one. However, it is a much bigger and more comprehensive exercise in performance reporting covering a wide range of services delivered by Australia's constituent units. They amount to almost \$150 billion covering over two-thirds of total government recurrent expenditure. Banks and McDonald describe how RoGS evolved — when they started in 1995 they already embraced a range of different public services (including education, health and justice). The 2011 *Report* contains performance information for 24 'overarching' services, encompassing 12 specific services. Describing the machinery of this vast exercise, including the intergovernmental context, the authors conclude that RoGS can overall be seen as a success in that it contributes to stimulate decision making and that it also reflects the cooperative nature of Australian federalism. However, they also see room for further improvement.

In Chapter 10, *Ben Rimmer* focuses on the COAG reform agenda from the perspective of the Commonwealth government and how it could potentially transform or remodel Australian federalism. COAG — the Council of Australian Governments — is the prime body dealing with dialogue, disputes and funding arrangements. At the core of the COAG reform agenda is the objective, notes *Rimmer*, to improve service delivery through three related provisions: funding linked to the achievement of outcomes and outputs (rather than inputs) in areas of policy collaboration; devolution of decision making and service design to the frontline wherever possible and effective; and competitive tensions between the constituent units ('competitive federalism') and competitive tensions between service providers. The use of benchmarking to measure performance underpins

these three elements, forming a cornerstone of the COAG reform agenda. The success, so far, of this ambitious agenda has been modest, argues *Rimmer*, since the Australian states that have not yet fully delivered on improving public services while at the same time receiving increased funding from the Commonwealth government.

An insider view on the machinery of the COAG Reform Council (CRC) is provided by *Mary Ann O’Loughlin* in Chapter 11. The CRC is the independent benchmarking assessor in Australian federalism’s new performance regime. She explains how more than 90 different payments from the Commonwealth to the States for specific purposes were combined into five new National Specific Purpose Payments. These are underpinned by National Agreements, concluded between the Commonwealth and the States, on key service delivery sectors of schools, skills and workforce development, health care, affordable housing, and disability services. Noting that the whole exercise is still in its early stages, *O’Loughlin* points out that there are some technical challenges CRC is seeking to overcome in cooperation with the States — notably regarding the conceptual adequacy of indicators as well as the availability of adequate data for reporting progress. She also mentions the issue of causality: ‘A comparative analysis does not explain why there are differences between the jurisdictions or why performance has improved or declined over time.’ However, CRC remains committed to pressure governments to take action in response to performance feedback in order to improve service delivery.

In Chapter 12, *Peter Dawkins* and *Sara Glover* discuss the National Agreements on Education in more detail from the perspective of the State of Victoria. They explain how through this agreement benchmarking has become firmly embedded in national policy through the setting of incentives and rewards for States and Territories. This involves payments from the Commonwealth linked with an outcomes framework. Their analysis concludes that there is a very important role for benchmarking in seeking to improve the educational system. However, they also identify a wide range of challenges in undertaking successful benchmarking and — agreeing with *O’Loughlin* — they emphasise the need to take account of the different contexts: when seeking to improve educational outcomes with the assistance of benchmarking, it is important to develop an understanding of what causes improvements in outcomes. Over time *Dawkins* and *Glover* expect that there will be a learning experience and that significant progress can be made to improve educational outcomes in Australia.

A view from Queensland on the value of benchmarking is provided in Chapter 13 by *Sharon Bailey* and *Ken Smith*. They analyse four benchmarking exercises that Queensland is involved in with the Commonwealth. Their main argument is that in employing benchmarking exercises, the issue of context needs to be taken very

seriously — an argument that is echoed also by other authors in this volume. Otherwise, so they claim, the opportunities for policy learning are lost. Describing how in 2010 Queensland had been penalised in the context of one of the Commonwealth–State benchmarking operations, they argue that sanctions might have an adverse impact on the relationship between States and the Commonwealth and hence on performance in the future. According to them, sanctions are a strong, but potentially blunt, tool that requires supplementation. Benchmarking, they contend, has the potential to be misused or to bring about unintended consequences; its impact is dependent on context and the way it is used.

In Chapter 14, *Helen Silver* emphasises the merits of the collaborative nature of intergovernmental relations in Australia. Good processes, so she argues, lead to good outcomes. Given that a decisive feature of federalism in Australia is its vertical fiscal imbalance, she points out that benchmarking exercises need to be shared exercises between the States and the Commonwealth. In this vein, she regards the ongoing institutional reform of COAG itself as important and argues for the need of an intergovernmental agreement to enshrine COAG’s principles and governance. This reform should entail some basic procedural disciplines, such as planning for a small number of regular meetings each year. As part of such an agreement, COAG should be provided with an independent secretariat to coordinate a more focused agenda and allow for the States and Territories to put issues on the table for discussion and action.

Outlook

This volume is attempting to draw out some preliminary comparative conclusions about the relation between ‘federalism’ and ‘benchmarking’. This is an ambitious exercise, and we realise, that this volume can only be a first step in a larger project. There needs to be more empirical research to determine the exact nature of that relationship — and to what extent benchmarking ‘delivers’.

Australia has been at the forefront of experimentation with benchmarking as a tool to improve policy performance. Commonwealth and State governments have, across a wide range of policy sectors, negotiated intergovernmental agreements that identify outcomes, goals, targets or guidelines, and include obligations on the part of participating governments to report to the public on the achievement of these measures. This form of benchmarking is attempting to avoid (or at least alleviate) the hierarchical and prescriptive relationship between Commonwealth and State governments associated with the traditional conditional grant programs that characterised Commonwealth–State fiscal arrangements prior to the ‘new public management’ era.

Also in other federal systems central governments and constituent units have to balance the centripetal and centrifugal impulses for country-wide policy outcomes, on the one hand, and policy outcomes that respect state autonomy or at least promote flexibility, on the other. The contributions by the international authors show that other federations are embarking on a similar route as Australia, without employing or copying its very comprehensive and systematic approach to public service benchmarking.

All chapters in this volume show that benchmarking has become an important aspect of federal governance and we believe that there is much that we can learn about inter-governmental benchmarking by looking across different federal systems. However, the contributions also show that benchmarking comes in many forms and that there are different drivers for benchmarking regimes. In practice, benchmarking is used to describe a wide variety of arrangements, and the objectives vary. While some are about accountability and transparency in intergovernmental relations and public service delivery, some focus on learning and improvement, and sometimes it is a mixture of both.

We also need more research on the impact of politics in general and the broader institutional context (including fiscal arrangements) on benchmarking regimes. Benchmarking can be viewed as an instrument of governance, but the issue of how to set up the governance of benchmarking regimes is also emerging as a key issue from many of the contributions of this volume requiring further investigation. One preliminary conclusion is that the models of a collegial nature, that are not based on hierarchy, targets and reputation effects (naming and shaming), encourage the greatest willingness of constituent units to participate. However, the jury stands out whether it is those arrangements that lead to performance improvement.

Another preliminary conclusion is that all benchmarking systems seem to face considerable challenges in creating and capturing robust and comparable indicator data. Producing good comparative data is only one step in the benchmarking process, and an equally important step is to ensure adequate analysis and interpretation of those data.

Many authors in this volume argue that benchmarking of public services matter because it is critical for governments and communities who need to know whether services are effective, efficient, who is accountable for service delivery, and whether the outcomes of service delivery are in the interests of the citizenry. They also argue that it is an important framework for policy decision-making. However, we would need to more research to find out what works, for what purposes, and with what opportunity costs.

And finally: What is the role of citizens and service users in benchmarking? Although the political rhetoric surrounding benchmarking and the putative benefits of federalism makes some assumptions about improved service delivery, in practice citizens and service users are often only marginal participants in many benchmarking systems. They are rarely involved in discussions about what the indicators should be, what they mean, or what should be done in response to benchmarking results.

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.

2 From the improvement end of the telescope: benchmarking and accountability in UK local government

Clive Grace¹
Cardiff Business School

2.1 Introduction

In the summer of 2010 the UK's newly elected coalition government announced the abolition of the principal benchmarking and performance management regime for local government in England, the Comprehensive Area Assessment, and its intention to abolish the principal authors and stewards of that regime, the Audit Commission (DCLG 2011) as well. The government also announced new requirements for public services to publish more information so that an 'armchair auditors' would be sufficiently equipped to hold those services to account directly, and without the intervening agency of bodies such as the Audit Commission.

These policies were introduced in the context of the wider programme of the coalition government with its emphasis on 'localism' and on the 'Big Society', and a comprehensive assault on the many intermediary and 'arms length bodies' such as the Audit Commission which were seen as fogging the relationship between government and citizenry. They were also a reaction to a decade or more of what was seen as top down performance management, inhibiting the exercise of professional discretion at the front line and creating a bureaucratic morass in which

¹ Clive Grace is currently an Honorary Research Fellow at Cardiff Business School and, previously, was a UK local authority CEO, and Director General of the Audit Commission in Wales. He was also Deputy Chair of the Beacons Scheme from 2003 to 2007. In consequence he had direct involvement and responsibility in respect of a number of the benchmarking and related programmes discussed in this piece and, accordingly, more than the usual caveats apply to the judgements expressed therein.

performance targets distorted behaviour and generated gaming as much as they generated genuine service improvement.

Yet the simultaneous celebration of the new ‘armchair auditors’ and sources of performance benchmarking information on local government services on which they might draw, such as that compiled by the Association of Public Service Excellence, gave more than a hint that the benchmarking era was changing rather than concluding. Indeed, it underlined the ubiquitous role and use of benchmarking in UK local government and other public services.

There are many varieties. First, there is a wide range of **service-based cost and technical comparisons** conducted as benchmarking ‘clubs’ of one kind or another. These include those of the Association of Public Service Excellence (APSE), a not-for-profit voluntary body established with service comparisons of ‘blue collar’ local government services as a core aim; the Chartered Institute for Public Finance and Accountancy (CIPFA), a major professional accountancy body for, *inter alia*, local government finance staff; and the Wales Audit Office (WAO), the statutory public audit body for Wales. Also in this area are the ‘communities of practice’ established across a range of different services by the Improvement and Development Agency (IDIA), an agency of the Local Government Association (LGA), which has now been absorbed within the LGA.

Secondly, there have been a series of centrally determined **performance indicator sets** with results often published in the form of league tables. Then there have been **performance regimes** for local authorities, looking at the whole organisation and testing them against pre-set frameworks, including the Comprehensive Performance Assessment (England), the Wales Programme for Improvement, and Best Value Audits (Scotland). These led in England to a yet wider programme of Comprehensive Area Assessments, which brought together data on a much wider group of local services. Alongside these performance regimes has been a programme of ‘voluntary’ assessments using **external peer review** methods against a framework underpinned by the European Framework for Quality Management. There have also been major **excellence benchmarking schemes**, which test projects and services against a pre-designed benchmark to identify best and excellent practice, and most notably the central government run Beacon Council Scheme in England.

These were (and still are, in some cases) all major programmes of work, and it all adds up to an awful lot of benchmarking, and to considerable direct and indirect cost. Perhaps unsurprisingly, therefore, it all figured strongly in the unfolding dynamic between local and central government amongst the constant striving for

improved local government services and performance that started under Thatcher but which gathered even greater pace under the Blair–Brown Labour governments.

There are good (structural) reasons why benchmarking in UK local government has been so widespread. Accordingly, **Section Two** briefly sketches out key features of UK local government along with an indication of the position in respect of the devolved governments of Wales and Scotland, and then describes the range and scope of UK local government benchmarking and why it matters. **Section Three** shows how the popularity and use of the various benchmarking methods has ebbed and flowed across the period of 1998 to 2010 as an expression of three distinct eras of local–central relationships, all of which were (and are) concerned, broadly, to achieve improved services and stronger accountability. Each of these eras has associated with it not only particular instruments to secure that improvement but also a ‘theory’ of improvement — sometimes explicit and sometimes implicit, and more or less comprehensive and credible — as to how those instruments might help achieve the desired outcome.

In **Section Four** we look in more detail at a selection of the benchmarking methods that have been deployed in and across these different eras — including such assessment as is available of the efficacy of the various rewards and punishments that they have employed to try and secure better services. In conclusion, **Section Five** draws out the underlying fundamental relationship that connects benchmarking to service improvement, and suggests a rudimentary framework through which policy makers and practitioners should approach the use of benchmarking methods. It also expresses some tentative views on the application of that framework to securing improvement in public services in an age of austerity, that being firmly the context of public services and their improvement in the UK for the foreseeable future.

In doing all this, the aim of the chapter is threefold. First it provides its own ‘benchmark’ of comparison as between the decidedly unitary governmental system of the UK and the divided governance characteristic of federal systems (see Fenna, this volume). This is not about ‘learning lessons from the UK’, but simply about providing a broader landscape of contrast and difference to help generate insight and understanding. Secondly, the creation of devolved governments covering the 13 per cent of the UK population in Scotland and Wales adds further to the comparative mix. It does not of course make the UK ‘federal’ in any fundamental sense. As yet, neither the devolved governments nor indeed any of the local government units in any part of the UK constitute truly different ‘orders of government’, each with their own separate source of constitutional validity and powers. But UK devolution has created lines of policy tension and divergence that can augment the value of UK-related comparison. Thirdly, there has been so much

performance benchmarking of UK local government, and quite a lot of academic study of its character and consequences, that it ought to be a source of some general propositions of value not only for the UK system itself but for other jurisdictions as well.

2.2 Features of UK local government and why benchmarking matters

The significance of benchmarking in the UK arises out of particular features of UK governance, including the size and scale of local government; its considerable service delivery responsibilities; and its financial dependence on central government.

Vertical fiscal imbalance is a core feature of UK public services, and that drives a great deal of benchmarking behaviour.

Structure and scale

Many areas have one level of principal local authority ('unitaries') and compared to most other jurisdictions they cover relatively large populations, typically between 100 000 and 1 000 000+ residents. This includes most major urban conglomerations, all of Wales and Scotland, and a growing number of largely rural areas. Some county areas retain two tiers of principal local authority — each county authority then contains a number of separately elected district councils. Typically the principal local authority in an area (or the county and districts combined in the two tier areas) will have expenditure responsibilities in the order of £3000 to £4000 per capita. So a unitary authority of 100 000 persons would have a revenue budget of between £300 million and £400 million per annum. If they provide all or most services in-house, the associated staffing requirement will be about 4000 full-time equivalents. About 80 per cent of their expenditure is met by grant from central government, allocated on a needs-based formula. The balance comes from local property taxes (the 'council tax') and from fees and charges.

Tasks

The principal local authorities are responsible for a very wide range of services. These include pre-school provision; primary and secondary education to age 18; and social services and social care for the vulnerable. Local authorities provide social housing either directly or via associated providers. They also have functions for

roads and for transport; economic development; and local regulation. Moreover, since 2000 the principal local authorities have an explicit and a growing role for ‘community leadership’. The precise form varies but this places responsibility on these authorities to plan comprehensively for their areas — including coordination of other local public services (police, fire and emergency, health, and so on) in jointly tackling issues that cross service boundaries such as community safety, or address the needs of groups requiring active collaboration between public services such as the elderly. More recently, the community leadership ambition has extended even further to encompass the role and impact of centrally-run services that are delivered locally, including skills training for example, initiated through a programme called ‘Total Place’ (though since re-labelled by the 2010 coalition government as ‘Community Budgets’).

Central–local relations

Unsurprisingly, given this landscape, central–local relations have long been a significant feature of policy and administration in the UK. They have always been dynamic, and under a degree of tension. In the Thatcher years the focus was on forcing local authorities to accept a degree of marketisation of services, and subjecting them to explicit performance measurement through the agency of the newly created Audit Commission. This continued under the Blair–Brown governments, which further developed an explicit and muscular centrally-run performance regime via ‘arms length’ bodies such as the Audit Commission. The working assumption by central government throughout has been that local government needed to improve and change — although as we shall see in Section Two, that assumption evolved across the period 1998 to 2010.

During the early years of the Blair–Brown governments, both Wales and Scotland gained a significant measure of devolved authority, and thereby gained responsibility for local government generally and for virtually all local government services. For these devolved governments central–local relations figure even more strongly, if only because local government in both nations is responsible for a significantly larger share (about 40 per cent) of their overall expenditure, all of which is financed from the UK level by way of grant aid to the devolved governments (Midwinter 2002). Relationships between central and local governments have tended to be closer in Scotland and Wales than in England. Nonetheless, the devolved governments have also operated performance regimes for local government, again via arms-length bodies including the functional equivalents of the Audit Commission — for example the Wales Audit Office (WAG 2005).

Local government performance and benchmarking

In the UK, local government performance matters well beyond local government itself principally because of the sheer cost and scale of the services they provide, and the level and nature of centrally provided finance and the concomitant limitations and constraints on local fund raising. UK local government effectively inverts the old adage about ‘no taxation without representation’ in that local communities enjoy ‘local representation (and service delivery) without local taxation (even the locally raised and administered council tax has been effectively capped by central diktat)’. For the very most part, local populations do not finance local services other than via the taxes they pay to central government, and there is only an imperfect and not widely understood relationship between the level of local council tax levied and the extent and quality of services provided. For this reason, coupled with the impact of national political sentiment on local elections which are generally fought under mainstream political party banners, local elections provide limited mechanisms for accountability of spending and of service performance.

At the same time, that performance is critical not only to the communities to whom the services are being delivered, but also to many national politicians and government ministers. National programmes are often delivered through and by local authorities, and so national politicians are keen to see them do well and to be able to claim the credit. Local politicians may be just as similarly well motivated but they have in addition the responsibility for actual delivery, with all its attendant problems.

These features of UK local government that prompt extensive performance measurement and associated benchmarking both exemplify and amplify wider tendencies in UK public services. Pollitt (2006) identifies the UK as an outlier in this respect, and traces the cause to a mixture of scale and centralisation, institutions, and culture, an analysis endorsed by Hood (2007) and James and Wilson (2010). The consequences in relation to benchmarking and local government in the UK have been profound. Benchmarking in all its many guises has become ubiquitous, and in so doing perhaps also reached the tipping point at which a ‘logic of escalation’ sets in (Pollitt *et al.* 2010) through which initially few and simple measures become more numerous and comprehensive, and also become summative through league tables and/or targets. These then become linked to incentives and sanctions, with associated pressures for gaming. The measures become more complex, and harder for non-experts to understand, and ownership of the measures becomes more diffuse, all resulting in a decay of public trust. As we shall see, attempts were made in the UK to avert or reverse this logic across the period from 1998 to 2010, but the effort it took to do so is itself testimony to the

underlying ‘natural’ dynamic that Pollitt identifies in public service performance measurement systems.

2.3 Eras and theories of improvement in UK (mainly England) local government

The benchmarking of local government services can be undertaken as an exercise with intrinsic value in order to facilitate ordered comparison and dialogue between service providers, with no more than an eye to avoiding being a laggard in terms of cost or technical standards. But even at this basic service comparison level it is generally stimulated by thoughts of wider progress and learning. When initiated at the governmental level, and especially when initiated by central government with the intention that it should apply to local government, then more is usually afoot, or so it has been in the UK. Benchmarking has been at the heart of a number of attempts by central governments to improve UK local government and its services. Those eras had their prologue in the Thatcher years, and benchmarking had a considerable role then. This was the genesis of the Audit Commission, and the comparisons of service cost that drove the regime of Compulsory Competitive Tendering where local authorities were required to expose (mainly blue collar) services to market forces. It was also when a national set of Performance Indicators was established for local services, for which data were collected and published nationally (Boyne 2002). It gathered force and reached its apogee across 1998 and 2010.

As that momentum increased, it became associated much more explicitly in England for local government with the theories of improvement associated with each of the two eras into which the Blair–Brown years divide of 1998–2005, and 2006–2010. The first was an era of a **central drive for local improvement** and the second was an era of **local drive for self-improvement and self-regulation**. They in turn have given way through a change of central government to a radically different approach, which is best described as a **central drive for localism**. As we shall see, a drive for localism does not mean entirely abandoning a role for benchmarking, though its role is undoubtedly different and considerably reduced, at least in the short term.

1998–2005: Central drive for local improvement

In this period the underlying theory of improvement was that local government would not improve unless it were driven to do so by muscular performance measurement and performance management by the centre. As explicitly articulated

by one of the Blair government's most senior public services adviser, the aim was to drive public services, including local government, from 'awful to adequate' (Barber 2006).

The chosen policy instruments started with the statutory requirement that local authorities should achieve 'Best Value' across all of their service provision, this being the optimal mix of efficiency, quality and effectiveness. This statutory aspiration was linked explicitly to two other instruments. First, local authorities were required to collect data to inform a new set of Best Value Performance Indicators with the data being validated and published in league table form by the Audit Commission. Secondly, they were required to conduct Best Value Reviews of all their services on a rolling programme against a pre-set statutory framework of which comparison with others was a central tenet, with those Reviews also being externally validated by the Audit Commission and the results published (Ball *et al.* 2002).

The Best Value programme was quickly seen as having two major failings. First, the volume of the Reviews themselves was seen to be excessive and indigestible by local authorities and by the Audit Commission alike, and of generally poor quality and lacking the 'bite' of successful challenge to existing poor services. Secondly, it was considered to be insufficiently focused on the things that really matter in determining whether a local authority improved.

The Comprehensive Performance Assessment (CPA) was introduced to remedy these defects, and this became the star performer and the most brutal instrument in the UK local government benchmarking armoury. It involved a periodic, composite, external judgement of each local authority by the Audit Commission of its corporate capacity and leadership, and also its services, with the authority being publicly scored across a range from 'Excellent' to 'Poor'. Those judged to be 'Poor' were then subject to central intervention and support without any choice in the matter.

These new instruments called not only for considerable effort and investment by local authorities, but also significant additional capacity and expenditure by the Audit Commission and by a number of other inspectorates of services (Davis and Martin 2008; Downe 2008; Martin 2006). In due course, CPA evolved during this period into CPA version two, as the 'Harder Test', reflecting the active management and development of the performance regime.

Running alongside CPA were other instruments, including the Beacon Council Scheme, which offered financial and reputational rewards for excellence and which promoted learning between councils from the best practice thereby identified and celebrated (and on which see Section Four, below). This clearly reflected a very

different improvement ‘theory’, but the respective direct costs of CPA and of the Beacon Council Scheme reflect their respective ‘theoretical’ weight in the mix — the expenditure for the Beacons was less than 2 per cent of that for CPA.

2006–2010: Local drive for self-improvement and self-regulation

By 2005 the local government improvement picture had changed considerably. As judged by CPA scores, things had got better generally, and the number of ‘poor’ authorities had declined considerably (Grace and Martin 2008). The Government’s five-year local government strategy of 2005 and the Local Government White Paper of 2006 both signalled that they had responded to the challenges (‘burden’) of a top-down benchmarking approach by shifting from centrally-driven performance measurement to a more ‘voluntary’ (local) system. The new local performance framework remained robust, and aimed to bring together a revised and more outcome focussed set of national performance indicators (approximately 190), local area agreements (LAAs) and comprehensive area assessment (CAA) to provide a coordinated focus for improving services and quality of life for local people. In beginning the move from CPA to CAA, central government underlined the community leadership role of local authorities, and signalled a much bigger role for local government itself in the drive for *self-improvement* and self-regulation. The underlying theory of improvement was that local government was increasingly performing well, and that top-down performance measurement and management both *could* therefore be eased back and, indeed, *needed* to be eased back if further improvement were to be achieved (Barber 2006).

Alongside the shift to a more sector-led approach there was a reduction in the level of inspection, intensifying a trend that had begun to emerge somewhat earlier (Bundred and Grace 2008). At the same time, ‘voluntary’ improvement activity by the Local Government Association and the Improvement and Development Agency intensified and claimed a leadership role (De Groot 2006), focused around the ‘voluntary’ alternative to CPA in the form of the sector-led Local Government Improvement Programme (Jones 2006). Renewed programmatic statements emphasising the need for central government to ‘let go’ and for local government to ‘take responsibility and move beyond compliance’ were published during this period (De Groot 2008).

2010–201?: Central drive for localism

The clamour for ‘localism’ from central and local government alike during the latter years of the Blair–Brown governments could have been a bridge of continuity

across to the new coalition government of 2010. In the event there has been a much greater sense of rupture in that for *ideological* as well as for practical reasons the Coalition government believes that local government should have responsibility for its own improvement as part of a wider theory of the relationships between state, society and public services. ‘Localism’ has been articulated as a major theme, along with idea of the ‘Big Society’ (Tuddenham 2010). The abolition of CAA and of the Audit Commission reflected the new theme, with ‘armchair auditors’ taking centre stage instead of central institutions and frameworks.

It is too early to gauge with confidence how the new framework of ideas and policies will develop, but it is already clear that this new era will not be a stranger to benchmarking. Indeed, the armchair auditors will require consistent, relevant, validated and national benchmarking data to be able to do their job properly. The new ‘shock troops’ are the Efficiency and Reform Group of the Cabinet Office and benchmarking is one of the pillars of their programme (Collier 2010). Further, in specific relation to local government, APSE has already claimed Coalition government endorsement for its current voluntary scheme.

Theories of improvement and the devolved governments

In Scotland and Wales, the instruments of local government benchmarking have been rather different — although to international eyes the similarities may look more significant than the differences. The governments of both Scotland and Wales deployed variants of CPA — Best Value Audit and the Wales Programme for Improvement — that similarly reviewed their local authorities on a ‘whole-organisation’ basis using their principal arms-length public audit body. Both also contained a similar emphasis on the importance of corporate capacity and leadership, and both made provision for intervention by the centre where failure was found. But there were also important differences, notably in the degree to which the results were assembled into scores or published as league tables (Martin *et al.* 2010). Importantly, these differences were expressly associated with different theories of public services change and improvement — against the competitive character and the use of the court of professional and public opinion employed by CPA, it was education and persuasion that was stressed by BVA, and collaboration and consensus by WPI (Martin *et al.* 2010).

2.4 Varieties of benchmarking and UK local Government

Benchmarking in one form or another has featured in all the ‘theories of improvement’ as applied to UK local government. Moreover, each benchmarking instrument carries — at least potentially — a ‘sub-theory’, which helps explain (were it to be articulated) what behaviour it is hoping to stimulate or inhibit through its application. Such theories are not, of course, always made explicit, and if they are they may not be right about the behaviour predicted. Nor is it always the case that where a bundle of instruments are explicitly assembled the resulting composite ‘theory’ will be internally coherent or fully comprehensive. Just as UK governments have been vigorous in their use of benchmarking for local government, so have they also been fairly explicit about what they hoped to achieve and how — but they may not have got it right. There is a quite strong and definite relationship between benchmarking instruments and theories of improvement but it is not always easy to pin down in particular instances.

So this Section reviews a fairly small number of the range of benchmarking techniques applied to local government in the UK with a view to explaining a little more about how they worked and also to explore what seems to work and not work in terms of the types of indicator and the reward/punishment mechanisms deployed, including whether any perverse incentives have been a significant or minor problem. The techniques reviewed are service-based benchmarking clubs; national performance indicator sets; whole organisation assessments; and ‘excellence’ benchmarking.

Service-based benchmarking

There are three initiatives to cover here. First APSE’s benchmarking focuses on the value of services provided by direct labour in local authorities, and can be seen as having its genesis in a partly defensive approach. The data sets have operated from 1998 to the present, and now cover 200+ local authorities. The data sets are practitioner developed in what is entirely a voluntary scheme, and mainly cover ‘manual’ services such as building cleaning and maintenance; civic venues; culture, leisure and sport; education catering; highways maintenance; parks, open spaces and horticultural services; refuse collection; sports and leisure facility management; street cleansing; street lighting; transport operations and vehicle maintenance. The data sets are extensive; for example, there are 30 key performance indicators for building cleaning. APSE see their approach as valuable in the ‘localism’ debate, and celebrate its endorsement by the coalition government (see, generally, www.apse.org.uk).

A second scheme is run by CIPFA, the professional accountancy body centred on the public sector. Their scheme too is voluntary. It has a financial focus, and draws together data on children's services; police; environmental matters (crematoria, cemeteries, waste, etc); housing (rents, homelessness, etc); leisure; personnel matters; and planning. Some of the data sets are long standing (10 years), and some much more recent. Membership costs £600 per local authority per service 'club', and is variable between them (www.cipfa.org.uk) .

A third is operated by the WAO, for example in the policy area of waste collection and disposal. They work to set up benchmarking clubs with the Wales Local Government Association (WLGA) and Welsh Assembly Government (WAG). As the public audit body for Wales with both value for money and performance responsibilities as well as financial audit, the WAO's reasons for operating the clubs are partly to achieve an accurate cost modelling for the review of the national waste strategy and to provide a policy baseline. The benchmarking data allow for greater comparisons between authorities, supporting the sharing of best practice and bringing together service improvement and efficiencies. WAO also provides annual financial reports on the waste management undertaken by local authorities and informs an annual report (WLGA 2010) and a Report to the Wales County Surveyors Group (WAO 2008).

There appears to be little evaluation of the impact or effectiveness of any of these voluntary arrangements, although their continuity and growth clearly indicates that the members appear to get sufficient value to warrant continuing to pay the (albeit modest) annual subscriptions. Nor is it certain that they would be so popular if the wider frameworks of performance measurement were not in place. They are, however, highly consistent with an approach emphasising professional front line leadership of service improvement.

National performance indicator sets

The best example of an instrument in this area is the set of Best Value Performance Indicators (BVPIs) that operated in England from 1998 through 2006 before they gave way to a more outcome focussed National Indicator Set of PIs. The BVPIs consisted of some 200+ indicators across all frontline and corporate services. They were set centrally — albeit after extensive consultation — and were regularly revised and supplemented in light of experience. This had the virtue of ironing out problems of interpretation or data collection and validation, but weakened the continuity of the data sets. The BVPIs were operated by the Audit Commission, and at their height required some 287 pages of guidance to try and ensure that data were collected in a standardised and comparable form (Boyne 2002).

As an example, the BVPI Equality Standard was articulated as the ‘level of the Equality Standard for local government to which the Authority conforms in respect of gender, race and disability’. The purpose of the standard was to provide ‘a framework for delivering continuous improvement in relation to fair employment outcomes and equal access to services to which all Authorities should aspire’. Authorities were required to report the level they had reached according to a six-point scale. This ranged from Level 0 (‘The Authority has not adopted the Equality standard for Local Government’) through Levels 1, 2 and 3 — the latter being that the ‘Authority has completed the equality action planning process, set objectives and targets and established information and monitoring systems to assess progress’ in addition to adopting a comprehensive equality policy (Level 1) and engaging in an impact and needs assessment (Level 2). The highest level (Level 5) is that the authority has achieved targets, reviewed them and set new targets, and is seen as exemplary for its equality programme.

The BVPIs reflect Pollitt’s (2010) ‘logic of escalation’ — not so much in their number, form and character, but inasmuch as they required ever increasing amounts of guidance to enable them to be operated in the preferred manner, and then themselves were put to one side in favour of a revised set of measures which were intended to be more ‘outcome’ oriented. At the same time, however, there was a persistent trend — especially after 2006 — towards reducing the number of indicators which local authorities were required to collect, on grounds that the collection of the data represented a regulatory burden on local authorities.

Whole organisation assessments

The approach of whole-organisation assessment and benchmarking of UK local government had its genesis in the Blair government’s policy concept of ‘Best Value’ and the idea of reviewing specific local government services comprehensively on a rolling programme against a framework of four ‘Cs’ — challenge, compare, consult, and compete — externally validated by the Audit Commission (Ball 2002). That gave way in the face both of the indigestibility of the sheer volume of reviews even by a radically expanded Audit Commission capacity, and the recognition that individual service performance rested as much (especially in the long term) on broader corporate capacity and leadership as it did on the ‘internal’ features of a particular service. The result was a centrally driven programme applying to all local authorities and delivered by the Audit Commission known as CPA. This was the ‘whole-authority’ external assessment against a pre-set framework focussed on leadership/corporate capacity and on key services, leading to a public score and with intervention for low scores, at a programme cost of some £200 million per annum in direct costs.

In Wales and Scotland the very interventionist approach of CPA was explicitly rejected, but both jurisdictions similarly developed on from the initial Best Value methodology to their own versions of whole organisation assessment (Grace 2007; Martin *et al.* 2010). In all three jurisdictions the approach was strongly associated with the development of ‘holistic’ public services inspection (Bundred and Grace 2008).

Excellence benchmarking

Just as significantly, the local government summit body for England also established a ‘voluntary’ whole organisation assessment methodology in the Local Government Improvement programme (LGIP) which evaluated individual local authority performance against an European Framework of Quality Management (EFQM) related model. This had twelve benchmark criteria: forward looking; community leadership; vision; corporate effectiveness; a performance focus; people capacity; values; participatory; member/officer relationships; resource usage; improvement; and well regarded externally. Although voluntary, its underlying approach also received endorsement from government as an improvement method (Bowerman 2002). It used a peer review methodology with mixed review teams, including local politicians from other local authorities, at a cost of some £20 000 per review. Whilst the LGIP was partly a defensive alternative to the government’s Best Value and CPA methods it emphasised a philosophy of ‘improvement from within’ (De Groot 2006), and there is some evidence of positive change associated with the reviews conducted under the LGIP (Jones 2004; 2006).

It was CPA, however, that was widely credited with driving up local government performance between 2001–02 and 2005–06 as judged both by the scores and by the views of almost all the key stakeholders (Grace and Martin 2008; Laffin 2008), although academics have been critical of many of its aspects (Leach 2010; Andrews 2004; Boyne 2004; Martin *et al.* 2010; Jacobs and Goddard 2007; Maclean *et al.* 2007). The public have been less convinced that local authorities have improved (Grace and Martin 2008), although there are data to support a link between public satisfaction with local authority services and CPA scores (Bundred 2006; Bundred and Grace 2008; Ipsos MORI 2007), and the public overwhelmingly support independent external assessment of local authorities.

Certainly CPA appears to have been more effective than the equivalent in Wales (Andrews and Martin 2007), although it is less clear that it has had more impact than the Scottish equivalent (Downe *et al.* 2008).

Insofar as local government has focussed on excellence and best practice identification and exchange, that too has been centrally driven. The Beacon Council Scheme is the most prominent example, with its focus on identifying best practice against pre-set criteria and then spreading the learning to other authorities. Eventually it became a 10-year programme across 1999–2009, funded centrally but operated by a local government summit body, and overseen by an Independent Panel appointed by a government minister. It was initiated partly to offset the widespread impression that the Blair government regarded local authorities as universally in dire need of doing better, although in comparative financial terms it was only a fig leaf of ‘respect’. Originally an England-only programme, Wales eventually established something similar, but rather more ‘collegiate’ in style.

The Scheme invited bids against ten themes each year chosen (after consultation with local government) by the Panel and by government ministries. They were generally fairly complex service/policy themes — for example, child mental health services, regeneration, and asset management — and often had strong ‘local–central’ aspects. It was popular throughout its 10-year life, with typically 250+ applications a year across the 10 themes. It cost some £5 million per annum, mainly in incentive grants to winners, showcasing and knowledge transfer events, specific grants for projects of knowledge transfer such as mentoring of neighbouring local authorities, and administration. Winners were announced at an ‘Awards Dinner’ event, and kept secret till announced.

The Scheme was subject to specific evaluation, and those evaluations were positive overall. They found that participating authorities used the Beacons’ experience to support knowledge acquisition and learning networks, and to underpin knowledge co-creation, and also engaged in actively applying the knowledge acquired to service improvement (Rashman *et al.* 2006). There were also opportunities for ‘vertical’ as well as ‘horizontal’ learning to be taken, which were especially important when so many complex policy problems depend for their success on effective ‘central-to-local policy/delivery’ chains. Such and other benefits do, however, depend upon there being sufficient capacity in the ‘receiving’ local authority (Rashman and Radnor 2005). There is also a broader question about the relationship of Best Practice programmes to wider questions of innovation, highlighting both the growing importance of innovation (Hartley 2005) and doubts as to whether Best Practice approaches do in fact stimulate innovation as compared to ‘top down’ approaches such as CPA. Somewhat counter-intuitively, the latter appears to have the edge (Brannan *et al.* 2008).

A trigger and not a silver bullet

If this overview of benchmarking of UK local government appears somewhat equivocal in the judgement of its efficacy, and that of particular benchmarking instruments, that should not be a surprise. UK local authorities are large, complex, multi-functional organisations that operate in a diverse and turbulent policy environment. As both Hood (2007) and James and Wilson (2010) have observed, there is much still to learn and understand about how ‘performance by numbers’ works and can be made to work better. In contrast to those who see a way forward through focusing on the way benchmarking itself is done (Tillema 2010), the answer if there is one almost certainly lies in appreciating more fully the relationship between a benchmarking instrument and the context of its use (Hill 2006).

An essential part of those varying contexts is the central–local relationship itself at any given time. The range, even at one particular moment, can be considerable (Grace and Martin 2008) and available scenarios as to how that relationship might develop are themselves associated with the deployment of performance instruments within it, and can carry a strongly normative character. Thus, a relationship of ‘targets and terror’ (Coulson 2009; Hood 2005) carries both potential risks and rewards (Jackson 2005) and a potential regulatory burden but one which may pay dividends (Bundred and Grace 2008). Central and local government and regulators alike may remain wedded to such models when they have already passed their optimum effectiveness (De Groot 2008), and when central government needs to let go and local government needs to move beyond mere compliance. In contrast, an era of ‘cooperation and contract’ in central–local relations invites the use of both different instruments and different behaviour (Young 2005) — especially if the focus is switched to achieving desired outcomes rather than merely desirable outputs (Wimbush 2010; Fenna, this volume). If the other end of the spectrum is reached — one that may be characterised as a locally driven approach of ‘initiative and innovation’ — then the role of benchmarking is likely to look very different, and perhaps much less intensive (AC 2007; Moore 2005; Albury 2005). As seen by one of the high priests of public service change and improvement in the UK, it is really a question of whether, for example, the entity to be improved needs to move from ‘awful to adequate’ or is rather at the stage of going from ‘good to great’ (Barber 2006).

2.5 Benchmarking for improvement

In a system like the UK’s, benchmarking has to be part of a suite of improvement tools because its scale and character is significantly subject to the influence of

national government and the consequences of local–central relationships. It has to be seen as an ‘arrow in the quiver’, rather than a silver bullet. At the same time, though, in the UK local government setting benchmarking is subject to counter-productive political and relational influences. These are not aligned to and often undermine the key features that make benchmarking more effective in terms of agreed and stable definitions; reliable data sets and reporting; agreement on the meaning of differences; and using difference as a starting point for inquiry associated with focussed change.

This political context sometimes gives rise to odd policy inversions when it comes to using benchmarking instruments. For example, the devolved governments of Wales and Scotland now operate more centralised performance regimes for local government than does England under the new Coalition government, whereas hitherto they were considerably less so. Moreover, the move to a radically ‘localist’ approach in England may well generate problems with benchmarking that may be similar to those experienced in federal systems — such as ensuring data and definitional consistency and reliability; establishing authoritative and widely supported performance indicators; coping with the extra difficulty of trying to judge outcomes rather than only inputs and outputs; and so on.

A further major contemporary challenge in the UK context, and in many other jurisdictions, will be how to best deploy benchmarking in an age of austerity and the attendant cuts in public expenditure and retrenchment in public services. In the UK at any rate it is widely acknowledged that long-term expenditure reductions will have to draw on change at the tactical, transactional, and transformational levels. At the **tactical** level — tightening efficiency in existing services, shrinking eligibility, and so on — *financial* indicators look to be the most useful. For **transactional** change — improving systems using ‘lean’ methods or better technology, for example — *process* benchmarks are likely to be more relevant. But for **transformational** change — tackling the ‘wicked’ issues, for example, that cross organisational boundaries, or where services are being completely re-designed around customer needs — probably only *excellence* benchmarking will be of any use at all, at least at the initial, innovatory stage when the early adopters are struggling at the leading edge.

Either way, the lesson is fairly clear. It is essential to think about and to deploy a combination of benchmarking (and other) tools from the *improvement* end of the telescope. Adopting an outcome focus, policy makers need to ask themselves: what do you want to get better? What is the current context of change, and what are the key relationships and forces shaping that context? How do you think change will happen — what is your theory of improvement? What will be the role of benchmarking within that? And how best can you optimise that role? This will still

not fashion a silver bullet of change from the benchmarking tools at their disposal, but it will perhaps help to ensure that the triggers for improvement are more likely to work in the right place and in a timely manner.

References

- Albury, D. 2005, 'Fostering Innovation in Public Services', *Public Money & Management*, vol. 25, no. 1, pp. 51–56.
- Andrews, R. 2004, 'Analysing Deprivation and Local Authority Performance', *Public Money & Management*, vol. 24, no. 1, pp. 19–26.
- , Boyne, G.A., Law J. and Walker, R.M. 2005, 'External Constraints and Local Service Standards: the case of Comprehensive Performance Assessment in English Local Government', *Public Administration*, vol. 83, no. 3, pp. 639–56.
- , and Martin, S.J. 2007, 'Has Devolution Improved Public Services? An analysis of the comparative performance of local public services in England and Wales', *Public Money & Management*, vol. 27, no. 2, pp. 149–56.
- Andrews, R. and Martin, S.J. 2010, 'Regional Variations in Public Service Outcomes: the impact of policy divergence in England, Scotland and Wales', *Regional Studies*, vol. 44, no. 8, pp. 919–34.
- AC (Audit Commission) 2007, *Seeing the Light: innovation in public services*, London.
- Ball, A., Broadbent, J. and Moore, C. 2002, 'Best Value and the Control of Local Government' *Public Money & Management*, vol. 22, no. 2, pp. 9–16.
- Barber, M. 2007, *Instruction to Deliver: Tony Blair, the Public Services and the Challenge to Deliver*, Politicos, London.
- 2006, 'Success — it's all in the execution', in Hill, R. (ed.) *In the Matter of How: change and reform in 21st Century public services*, Solace Foundation Imprint, London.
- Bennett, C.J. 1997, 'Understanding Ripple Effects: the cross-national adoption of policy instruments for bureaucratic accountability', *Governance*, vol. 10, no. 3, pp. 213–233.
- Bowerman, M. 2002, 'Isomorphism without Legitimacy?', *Public Money & Management*, vol. 22, no. 2, pp. 47–52.
- Boyne G.A. 2003, 'Sources of Public Service Improvement: a critical review and research agenda', *Journal of Public Administration Research and Theory*, vol. 13, no. 3, pp. 367–94.

-
- 2002, ‘Concepts and Indicators of Local Authority Performance’, *Public Money and Management*, vol. 22, no. 2, pp. 17–24.
- and Enticott, G. 2004, ‘Are the “Poor” Different?’, *Public Money & Management* vol. 24, no. 1, pp. 11–18.
- Bradbury, J. 2003, ‘The Political Dynamics of Sub-State Regionalisation: a neo-functional perspective and the case of devolution in the UK’, *British Journal of Politics and International Relations*, vol. 5, no. 4, pp. 543–75.
- Brannan, T., Durose, C., John, P. & Wolman, H. 2008, ‘Assessing Best Practice as a Means of Innovation’, *Local Government Studies*, vol. 34, no. 1, pp. 23–38.
- Broadbent, J. & Laughlin, R. 1997, ‘Evaluating the “New Public Management” reforms in the UK: a constitutional possibility’, *Public Administration*, vol. 75, no. 3, pp. 487–507.
- Bundred, S. 2006, ‘The Future of Regulation in the Public Sector’, *Public Money & Management*, vol. 26, no. 3, pp. 180–88.
- and Grace, C. 2008, ‘Holistic Public Services Inspection’, in Davis, H. and Martin, S.J. (eds), *Public Services Inspection*, Jessica Kingsley, London.
- Clarke, J., Gewirtz, S., Hughes, G. and Humphrey, J. 2000, ‘Guarding the Public Interest? Auditing public services’, in Clarke, J., Gewirtz, S. & McLaughlin, E. (eds), *New Managerialism, New Welfare?* SAGE, London.
- Collier, S. 2010, ‘The Work Programme of the Efficiency and Reform Group’, presentation to the *Efficiency in Central Government Conference*, 30th November 2010, London: Inside Government.
- Coulson, A. 2009, ‘Targets and Terror: government by performance indicators’, *Local Government Studies*, vol. 35, no. 2, pp. 271–81.
- Davis, H. and Martin, S.J. (eds) 2008, *Public Services Inspection*, Jessica Kingsley, London.
- DCLG (2011) *Future of Local Public Audit: consultation*, Department for Communities and Local Government, London.
- De Groot, L. 2006, ‘Generating Improvement from Within’, in Martin, S.J. (ed.) *Public Service Improvement*, Routledge, Abingdon.
- and Rogers, P. (eds) 2008, *Getting so Much Better all the Time*, London: Solace Foundation Imprint.
- 2008, ‘Challenges to the Approach to Improvement’, in Grace, C. and Martin, S.J. (eds) *Getting Better all the Time?*, IDEA, London.

-
- Downe, J.D., Grace, C.L., Martin, S.J. and Nutley, S.M. 2007, 'Comparing for Performance Regimes in England, Scotland and Wales', *ESRC Public Services Programme Discussion Paper 0705*, www.publicservices.ac.uk/library.
- 2008, 'Inspection of Local Government Services', in Davis, H.D. and Martin, S.J. (eds), *Public Services Inspection*, Jessica Kingsley, London.
- , Grace, C.L., Martin, S.J. and Nutley, S.M. 2008, 'Best Value Audits in Scotland: winning without scoring?', *Public Money & Management*, vol. 28, no. 2, pp. 77–84.
- Grace, C. (ed.) 2007, *Comparing for Improvement*, Solace Foundation Imprint, London.
- and Martin, S.J. (eds) 2008, *Getting Better all the Time?*, IDeA, London.
- Hartley, J. 2005, 'Innovation in Government and Public Services', *Public Money & Management*, vol. 25, no. 1, pp. 27–34.
- Hill, R. (ed.) 2006, *In the Matter of How: change and reform in 21st Century public services*, Solace Foundation Imprint, London.

PART I

INTERNATIONAL CONTRIBUTIONS

3 The implementation of the *No Child Left Behind Act*: toward performance-based federalism in US education policy

Kenneth K. Wong¹
Brown University

3.1 Introduction

In his last public policy speech at the end of his two-term presidency, George W. Bush chose to highlight his accomplishments in education reform in the General Philip Kearny School in Philadelphia. Seven years after the passage of the *No Child Left Behind Act* (NCLB), the President claimed that ‘fewer students are falling behind’ and ‘more students are achieving high standards’ (Eggen and Glod 2009). To address the concern that testing is punitive, the President commented, ‘How can you possibly determine whether a child can read at grade level if you don’t test? To me, measurement is the gateway to true reform’. This debate over the benefits and limitations of the federal Act continues as the Barack Obama administration begins to work on the reauthorisation of legislation. President Obama supports the federal role to strengthen accountability, including annual student testing. At the same time, he sees the need to provide additional resources to schools so they can improve teacher quality and improve student readiness for post-secondary opportunities.

Federal assertiveness in NCLB is a significant departure from a long held tradition of federal permissiveness. The federal government has been mindful that States have the constitutional authority over public education and that the American public adheres to the ethos of local control over public schools. The United States, in essence, maintains a decentralised education system. With the enactment of the federal NCLB, however, a more activist federal government seems to have emerged. To be sure, it is clearly not ‘nationalisation’ of education since there is no

¹ Kenneth K. Wong is the Annenberg Professor and Chair, Education Department, Brown University, Providence, USA.

national examination and States continue to design their own academic standards and decide on how to intervene in persistently under performing schools. Nonetheless, the federal law requires all States to apply federal criteria in holding schools accountable. A central feature is the requirement that students in selected grade levels must be tested annually in reading and mathematics and that the results be used as evidence to meet the academic proficiency targets as established by the Adequate Yearly Progress (AYP) to avoid federally-defined interventions.

Given the fundamental shift in the role of federal government following the enactment of NCLB, this paper examines the policy's institutional context, programmatic design, and implementation lessons. Several issues will be addressed. First, how does NCLB depart from the traditional federal role? Second, what are the key benchmarks on academic proficiency? Third, how does federal benchmarking reconcile with a decentralised system of education? Fourth, to what extent are State and Local governments responding to the NCLB sanctions and incentives, including growth in charter schools, contracting services, and restructuring low-performing schools? Finally, what are the implications for sustaining performance-based federalism?

3.2 From dual federalism to categorical federalism

Historically, the U.S. federal government has taken a permissive role in education that is consistent with what political scientist Morton Grodzins characterised as 'layer cake' federalism. Article I, Section 8 of the U.S. Constitution specifies the 'enumerated powers' that Congress enjoys and the Tenth Amendment granted State autonomy in virtually all domestic affairs, including education. Sovereignty for the States was not dependent on the federal government but instead came from the State's citizenry. Consistent with this view, in *The Federalist Papers*, published during 1787 and 1788, James Madison suggested a line of demarcation between the federal government and the States. In *Federalist* no. 46, he wrote, 'The federal and State governments are in fact but different agents and trustees for the people, constituted with different powers, and designed for different purposes'. The dual structure was further maintained by local customs, practice, and belief. It came as no surprise that in his description of the American democracy in the mid-nineteenth century, Alexis de Tocqueville opened his seminal treatise by referring to the local government's 'rights of individuality'. Observing State–Local relations in the New England townships, de Tocqueville (2000, p. 63) wrote, 'Thus it is true that the tax is voted by the legislature, but it is the township that apportions and collects it; the existence of a school is imposed, but the township builds it, pays for it, and directs it'. Public education was primarily an obligation internal to the States. The division of power within the federal system was so strong that it continued to preserve State

control over its internal affairs, including the *de jure* segregation of schools, many decades following the Civil War.

Federal involvement

Federal involvement in education sharply increased during the ‘Great Society’ era of the 1960s and the 1970s. Several events converged to shift the federal role from permissiveness to engagement. With the conclusion of the Second World War, Congress enacted the ‘G.I. Bill’ to enable veterans to receive a college education of their choice. Cold War competition saw the passage of the *National Defense Education Act* in 1958 — shortly after the Soviet Union’s satellite, Sputnik, successfully orbited the earth. At the same time, the 1954 landmark Supreme Court ruling on *Brown v. Board of Education* and the Congressional enactment of the 1964 *Civil Rights Act* sharpened the federal attention to the needs of disadvantaged students. Consequently, the federal government adopted a major antipoverty education program in 1965, Title I of the 1965 *Elementary and Secondary Education Act* (ESEA).

The ESEA, arguably the most important federal program in public schools in the last four decades, signalled the end of dual federalism and strengthened the notion of ‘marble cake’ federalism where the national and sub-national governments share responsibilities in the domestic arena. Prior to the 1965 law, there was political deadlock on the role of federal government in Congress. The States outside of the South were opposed to allocating federal funds to racially segregated school systems. Whereas some lawmakers refused to aid religious schools, others wanted to preserve local autonomy from federal regulations. Political stalemates were reinforced through bargaining behind closed doors among the few powerful Congressional committee chairmen (Sundquist 1968).

The federal role

The eventual passage of ESEA and other social programs marked the creation of a complex intergovernmental policy system (see also Fenna, this volume). To avoid centralisation of administrative power at the national level, Congress increased its intergovernmental transfers to finance State and Local activities. During the presidency of Lyndon B. Johnson, categorical (or specific purpose — see Fenna, this volume) programs, including Title I, grew from 160 to 380. By the end of the Carter administration in 1980, there were approximately 500 federally-funded categorical programs. Particularly important was the redistributive focus of many of these categorical programs that were designed to promote racial desegregation,

protect the educational rights of the learning disabled, assist English language learners, and provide supplemental resources to children from at-risk backgrounds. Despite several revisions and extension, ESEA Title I, for example, continues to adhere to its original intent ‘to provide financial assistance ... to local educational agencies serving areas with concentrations of children from low-income families to expand and improve their educational programs . . . which contribute particularly to meeting the special educational needs of educationally deprived children’ (ESEA of 1965).

Federal engagement in redistributive policy is reflected in its spending priorities. As suggested in table 1, federal contribution accounted for 8.5 per cent of the total revenues for public elementary and secondary education during 2006-07, a noticeable increase from 6.6 per cent in 1995-96. This increase occurred at a time when per pupil total spending rose from \$8833 to \$11 941 in real terms.

More importantly, growth in federal aid continues to associate with the policy focus on disadvantaged populations. As suggested in table 2, federal aid to programs for special-needs students showed persistent growth in real dollar terms. Between 1996 and 2005, these programs amounted to over 60 per cent of the total federal spending in elementary and secondary schools. The Title I program for the education for the disadvantaged increased from \$8.9 billion to \$14.6 billion in 2005 constant dollars. Federal aid in special education more than doubled, while the school lunch program increased its funding from \$9.8 billion in 1996 to \$12.2 billion in 2005. Head Start also jumped by 50 per cent in real dollar terms during this period. This trend of growing federal involvement in programs for the disadvantaged continues in the Obama Administration (see table 7 and the discussion later in the paper).

Redistributive federal grants have also taken on several institutional characteristics. First, under the grants-in-aid arrangement, the federal government provides the funds and sets the programmatic direction, but the delivery of services is up to State and Local agencies. Second, categorical grants focus on well-defined eligible students and only they would receive the services. Third, non-supplanting guidelines have ensured that federal resources are not diverted away from the eligible beneficiaries. Fourth, because categorical grants are widely distributed to schools and districts across many Congressional districts, bipartisan political support remains strong for these programs.

Money without results

As public schools show mixed performance, policy makers become growingly concerned about the effectiveness of federal grants. The passage of *Improving*

America's Schools Act (IASA) during the Clinton administration in 1994 signalled the beginning of federal efforts to improve program coordination and to allow for public school choice and competition. Among the most important features in the IASA was a provision that encouraged State and Local education agencies to coordinate resources in schools with high percentage of children who fell below the poverty line. The 'school wide' initiative was designed to phase out local practices that isolated low-income students from their peers in order to comply with the federal auditing requirement on the 'supplement nor supplant' guideline. Further, the IASA enabled public school competition with the allocation of federal charter school start up planning grants.

The IASA also aimed at monitoring schools that persistently failed to meet State proficiency standards. However, the legislation did not specify the consequences when schools repeatedly fell short of the federal expectations. IASA required States to adopt standards aligned with State assessments, but allowed them full autonomy to make instructional, governance, and fiscal policy decisions to support their academic performance standards. The political reality was that holding schools and districts accountable to high-stakes mandates was not feasible under IASA. There was very little enforcement of the IASA provisions and few States made substantial progress in meeting its requirements.

In short, categorical federalism takes a primary focus on the level of resources, regulatory safeguards, and other 'inputs' to meet the learning challenges of special-needs students. In providing supplemental funds to State and Local government, the federal government has not pressed for accountability in student achievement. However, with the *No Child Left Behind Act* (NCLB) of 2001, the federal government aims at combining an input-based framework with outcome-based accountability. In this regard, NCLB constitutes the latest evolution in our intergovernmental system in education in the United States.

3.3 Beyond categorical federalism: performance-based benchmarking in NCLB

The passage of the 2001 *No Child Left Behind Act* marked the beginning of a serious effort toward performance-based federalism. For some analysts, NCLB has changed the terms of federal–State relations to such an extent as to represent a 'regime change'. In his historical review of the federal role, McGuinn (2005) sees NCLB as a 'transformative' moment in that well-entrenched political interests depart from their traditional policy positions. Conservatives were ready to set aside their strong belief in local control and to endorse a visibly stronger federal presence

in education. Liberals moved to support a fairly comprehensive set of accountability measures, including annual testing of students in core subject areas with consequences. The de-alignment of traditional political relationships, as I might characterize the enactment of NCLB, embodies a fundamentally different set of ideas, interests, and institutions. Education now occupies centre stage in the political discourse at the national level.

The NCLB regime

With NCLB, all students and schools are required to meet Adequate Yearly Progress (AYP), a set of standards that are established through State-specified academic proficiency plans. All schools, including Title I schools, must test all of their students and report their test scores by racial, income, and other special need categories. More specifically, the 2001 law requires annual testing of students at the elementary grades in core subject areas, mandates the hiring of ‘highly qualified teachers’ in classrooms, and grants State and Local agencies substantial authority over failing schools. By linking the progress of schools and teachers to achieving a nationally specified rate of progress on State tests, these federal requirements aim at shaping curriculum and instruction in the classroom. In other words, federal mandates are no longer limited to schools that serve predominantly disadvantaged students as defined under categorical federalism. Instead, federal NCLB performance-based expectations apply to all students in all schools.

Focus on school-level and subgroup achievement

To determine if a school meets Adequate Yearly Progress annually in NCLB, student achievement is aggregated by grade and by subject area for each school. The School-level report includes the share of students proficient in each of the core content areas; student participation in testing; attendance rates; graduation rates; and dropout rates. Equally important, depending on their socio-economic characteristics, schools are required to report the academic proficiency of students in the following subgroups: economically disadvantaged students; students from major racial and ethnic groups; students with disabilities; and Limited English Proficiency (LEP) students. All students in grades 3 through 8 and an additional grade in high school are tested annually in mathematics, reading/language arts, and science.

In this regard, NCLB has made the achievement gap within a school more transparent for accountability purposes. By showing the percentage of students in each subgroup who attain proficiency in the academic content areas tested, the school must face the challenge of uneven distribution of academic outcomes. In

their study of Florida, Hall, Wiener, and Carey (2003) found that some schools that received an ‘A’ under the state’s own accountability system would no longer be labelled as successful. For example, the research team found an ‘A’ school where about 50 per cent of the students were proficient in reading and mathematics. The school was made up of 31 per cent white and 59 per cent black students. It had 57 per cent low-income students. The achievement data showed that while 90 per cent of the white students achieved proficiency in reading and mathematics, the proficiency rates for black and low-income students were only at 22 per cent in reading and 15 per cent in mathematics. In other words, the reporting requirements on subgroups have made the data on achievement gap accessible to the public.

Adequate yearly progress

NCLB allows each State to decide its own pace for reaching the federal goal of 100 per cent proficiency by 2014. Within this time frame, States are required to specify the annual targets on the percentage of students who meet the State proficiency standards in the core subject areas. These annual, measurable targets are known as ‘Adequate Yearly Progress’ (AYP). Districts, schools, and various subgroups are required to meet the AYP in order to avoid sanctions. According to the legislative intent of NCLB, States are required to identify a starting point of per cent proficient and establish the annual increments it will take for schools to reach proficiency. Schools are allowed to average the per cent proficient for each subgroup over two consecutive academic years, as well as over all grade levels tested in the school to calculate whether the school has made AYP.

NCLB uses a straightforward, absolute score cut-off in determining whether a school or a subgroup meets AYP. This ‘status’ approach has been criticised for its lack of consideration for academic gains that may not have resulted in reaching the proficiency cut-off standard. For example, a student who was over a year behind in reading could make achievement gains equalling a grade level of improvement and still be below proficient. In response to these concerns, the U.S. Department of Education has granted waivers (see Fenna, this volume) to several States to experiment with growth models for determining AYP (Olson and Hoff 2005). More recently, the Obama administration indicated its support for using academic growth as a measure for meeting AYP.

The goal of having the AYP is to connect current level of student achievement to the ultimate objective of reaching 100 per cent proficiency by 2014. The base line is set for 2002-03, where States were given the discretion to define their starting points in terms of the percentage of students who reached proficiency in the 2002-03 school year in language arts and mathematics. These starting points specify

how far schools have to go over the 12-year period in order to reach 100 per cent academic proficiency by 2014. Table 3 summarizes the starting point, the AYP, and the end goal as specified in a sample of State accountability plans. AYP targets vary by content area and grade level. However, these targets are applied to all students, including members of various subgroups. For example, Florida established the starting point for all grade levels at 30.68 per cent and 37.54 per cent for Language Arts and Math respectively. Michigan, on the other hand, has established separate starting points for elementary, middle, and high school grades in both language arts and math.

The pace of meeting AYP varies among States. For example, Arizona, Arkansas, Hawaii, and South Carolina, among others, have relatively low starting points compared to Colorado, Georgia, and Tennessee. The former would have to make greater progress on their standards in order to meet the 100 per cent proficiency target by 2014. States' accountability plans suggest three broad patterns to ensure that students reach proficiency: equal yearly goals; steady stair-step; and accelerating curve (Wong and Nicotera 2006). In the equal yearly goals approach, the AYP targets are set as equal increments every year until the 2014 deadline for 100 per cent proficiency. Annual equal increments are calculated by subtracting the starting point proficiency from 100 and then dividing by 12. The steady stair-step approach increases the AYP targets incrementally every two or three years to meet the 2014 deadline. In the third approach, States create an accelerating curve for improvement where the per cent of students meeting proficiency will rise slowly in the initial years but with greater gains occurring closer to the 2014 deadline.

The extent to which a district or a school meets AYP is affected by the number and size of students in the subgroups. In their analysis of this issue in California, Kim and Sunderman (2005) found that the percentage of schools meeting AYP declines as the number of subgroups rise in these schools. While 78 per cent of the schools with only one subgroup met the reading AYP in 2003, only 25 per cent of the schools with six subgroups were able to do so. When the authors considered the AYP data in Virginia, they found that 85 per cent of the schools that met both the State and federal proficiency standards had two or fewer subgroups. Only 15 per cent of the proficiency schools had three or more subgroups.

Because schools with a high concentration of subgroups (such as English Language Learners) may face greater difficulty in meeting AYP targets, the federal government allows schools to meet AYP by fulfilling a 'safe harbor' provision. Under this guideline, a subgroup is deemed as 'meeting' AYP if the percentage of students in the 'below basic proficiency' level is reduced by 10 per cent from the previous year. In Philadelphia, a large urban school district with 266 schools in Pennsylvania, for example, of the 158 schools that made the AYP in 2010, 37 per

cent of them met AYP by achieving the ‘safe harbor’ target. Nationwide, schools that fail to meet the State proficiency standards are becoming more transparent to the public and policymakers. As Table 4 suggests, during 2005, six States have the majority of the schools that did not meet the AYP. In another sixteen States, as many as 49 per cent of the schools failed the State proficiency standards.

Corrective actions directed at persistently low performing schools

Performance-based federalism is reinforced by federal threats and sanctions. NCLB calls for a set of ‘corrective actions’ when districts and schools fail to make AYP for consecutive years. The AYP applies not only to the overall performance of the school but also specific racial/ethnic and special needs subgroups within a school. Corrective actions and other sanctions, in other words, are aimed at closing the achievement gaps.

Federal sanctions intensify as schools and districts experience consecutive years of academic failure. These sanctions begin with the relatively modest requirement for a school improvement plan, options for families in schools not making adequate yearly progress to transfer to another public or charter school, and the implementation of supplemental educational or tutorial services after-school. In other words, sanctions in the first years of academic failure are not designed to change the school or district structure. Following four consecutive years of failure, NCLB allows for the more intensive sanctions. These include State-driven interventions that alter school governance and hiring decisions, such as school or district takeovers and replacement of personnel in poorly performing schools. It has been estimated that about 5000 schools in the US are eligible targets for the more drastic sanctions.

3.4 The challenge of implementing performance-based accountability

The emergence of performance-based federalism has created implementation challenges in the intergovernmental policy system. The federal government has relied primarily on State and Local capacity to implement the policy. Manna (2006) argues that ‘borrowing strength’ from State governments can facilitate federal capacity in the education policy arena where the social licence is historically weak. At the same time, tensions arise when many State and Local systems have limited capacity in analysing large scale data on student performance on an ongoing basis, in providing alternative instructional services in failing schools, and in making achievement and other schooling information more transparent to parents in a

timely manner. As Manna (2010) observes, the federal goal to promote accountability tends to conflict with the reality of school and district practices.

Pace of implementation shaped by states

During the initial implementation phase, States took incremental steps in meeting the federal legislative expectations. With dozens of states suffering from budgetary shortfalls in the 2000s, States delayed their response to seemingly costly federal mandates (on which, see Fenna, this volume). According to a 50-State report card on the first anniversary of the federal legislation, only five States received federal approval on their accountability plan (Education Commission of the States 2003). Further, only half of the States were prepared to monitor performance of various subgroups and to undertake corrective actions in failing schools. Over 80 per cent of the States were not ready to meet the federal expectation on placing highly qualified teachers in the classroom. It was only during the fourth year of NCLB that all the States had their accountability plans approved by the federal government. As Table 4 shows, in 2004-05, only 27 States (or 52.9 per cent) had at least 75 per cent of their schools meeting the federal AYP requirements. State capacity to meet AYP was seriously challenged as the proficiency cut-off level continued to rise toward the 100 per cent level on proficiency for all students.

Because States established their baseline that shaped AYP, academic proficiency was not uniformly defined. In an analysis of low-performing high schools, Balfanz and others (2007) found that schools in States with lower proficiency standards were more likely to make AYP. Schools in States with fewer subgroups for NCLB accountability purpose were also more likely to meet AYP. Further, the study found that some states focused on proficiency in 11th or 12th grade and graduate rates, thereby ignoring students who dropped out of the school system prior to the 11th grade testing time. In other words, variations in State academic standards made it difficult to assess the progress of NCLB across different states.

Resistance on testing requirements and accountability

Political opposition to NCLB arose in a number of States over the testing and accountability provisions (Wong and Sunderman 2007). First, States registered their opposition with legislative actions. In 2004, the Virginia House passed a resolution calling on Congress to exempt States such as Virginia, that had a well-developed accountability plan in place, from the NCLB requirements. The resolution called NCLB ‘the most sweeping federal intrusion into state and local control of education in the history of the United States, which egregiously violates the time-honoured

American principles of balanced federalism and respect for state and local prerogatives' (House Joint Resolution No. 192, passed January 23, 2004). The resolution passed 98 to 1, with the lone dissenter a Democrat. Further, after extensive lobbying by the Bush administration, the Republican controlled Utah House modified a law that would have prohibited that State from participating in NCLB. Instead, the law was amended to prohibit the State and Local districts from implementing NCLB unless there was adequate federal funding (H.B. 43 1st Sub, passed 10 February 2004). Other States — including Vermont, Hawaii, Connecticut, North Dakota, Oklahoma, and New Hampshire — passed similar resolutions. Indeed, during the first two years of NCLB implementation, the National Conference of State Legislatures identified 28 States that considered resolutions or bills requesting waivers, more flexibility and/or money, or that would prohibit the State from spending its own funds to comply with NCLB or even participating in the NCLB program. Moreover, in March 2004, the chief State school officers from fifteen States sent Education Secretary Rod Paige a letter asking for more flexibility in determining which schools were making adequate yearly progress.

Second legal action was taken against the federal government. The first legal challenge came from a coalition of districts in Michigan, Texas, and Vermont and the National Education Association, the nation's largest teachers' union. The plaintiffs argued that NCLB imposed federal mandates without adequate financial support. In November 2004, a federal judge in the US District Court for the Eastern District of Michigan rejected the challenge. The ruling stated that Congress had the authority to specify policy conditions on States (Janofsky 2005, p. A14). Subsequent rounds of appellate court decisions led to the U.S. Supreme Court, which decided in 2010 not to interfere with the earlier district court decision.

Another suit was filed by Connecticut against the U.S. Department of Education. That State sought full financial support from the federal government for the \$41 million State fund it spent to implement NCLB between 2002 and 2008 (Walsh 2010). The State also claimed that the federal agency had acted in an 'arbitrary and capricious manner' in deciding on State requests for waivers and exemption (Janofsky 2005, p. A14). For example, Connecticut cited that the Department of Education rejected the State's request for testing the students every other year instead of annually. The U.S. Court of Appeals dismissed the suit in July 2010 because the federal government had not taken any actions against Connecticut on NCLB implementation. The three-judge panel stipulated, however, that Connecticut could take administrative action against the federal government.

Limited implementation on choice and supplemental education services (SES)

Incremental steps were taken by States and districts on those NCLB provisions that aimed to reshape public education. Corrective actions provided a good example. A 2007 U.S. Department of Education study showed that participation in school choice and supplemental education services (SES) varied by grade level, with elementary school students having the highest participation rates. African American students had the highest participation rates in SES and above average participation rates in choice. Hispanic students had higher participation than Whites in SES, but lower participation rates than Whites in choice. Overall, participation in SES led to statistically significant improvement on math and reading scores, and spending more years in SES led to greater improvement in achievement. There was no significant effect on achievement from participating in the transfer option; however, the authors note that the small sample size of nine urban districts limited the power of this observation.

Case studies on the implementation of transfer options for student in low-performing schools generally found limited degree of local implementation. In his study of California, for example, Betts (2007) observed that school choice, as stipulated by NCLB, was largely underutilised throughout the State. In the initial year that schools were required to offer bussing for students who wish to transfer, some of this problem may be attributed to the timing of data being available to school districts, and, thus to parents. Additional reasons for limited implementation included the failure of districts to communicate clearly to parents the choice program, an inadequate number of places in better performing schools, and lack of parental interest in moving their children to schools outside of their neighbourhood.

The California case also showed that participation rates in SES were low, though not nearly as low as for the transfer option. Difficulties cited by State personnel with regard to implementing SES included a general lack of information about State-approved providers; districts were tardy in providing parents with information about SES; and some districts did not allow non-district providers to work on district property. Districts also had a considerable number of complaints/concerns about SES providers. Finally, like school choice, one of the greatest impediments to participation in SES was the substance and form of communications sent by districts to parents.

Cautions on school restructuring to turnaround low performance

While NCLB relied on State Education Agencies to implement the law's provisions, it did not pay adequate attention to their capacity to carry out the responsibility (Sunderman and Orfield 2006). After all, State intervention in low performing schools prior to NCLB was limited and not very effective (Mintrop 2004; Mintrop and Trujillo 2005; Sunderman and Orfield 2006). Not surprisingly, implementation studies found largely limited State and Local response to adopt the more drastic restructuring options. Even when schools persistently failed for consecutive years, States and districts were more likely to stay away from school turnarounds, where the principal and the majority of the staff would be replaced. Instead, most restructuring efforts aimed at extending the school day, bringing in outside expertise, and instituting modest changes to school governance. Indeed, implementation studies suggested that States generally narrowed the pool of schools that were targets for restructuring. For example, in Illinois, only 27 per cent of the districts that enrolled a substantial number of Title I students in 2004-05 were required to implement restructuring strategies, including personnel reassignment. In New York State, the number of Title I students receiving supplemental tutorial services grew from 31,700 in 2002-03 to 70 600 in 2005-06. In other words, the State took three years to increase from 13 per cent to 32 per cent of Title I students who were eligible for these services (Center on Innovation and Improvement 2006).

Need for organisational accommodation

Facing Local and State resistance, the U.S. Department of Education relaxed certain requirements on a case-by-case basis (Sunderman 2006; Hess and Petrilli 2006). Among the first policy changes the federal government made concerned the inclusion of students with disabilities and English language learners into the State accountability system. The policy shift was in response to State and Local objections to holding all students with disabilities to grade-level standards and the challenges of implementing the NCLB requirements for English language learners. States with a higher concentration of these two subgroups were more likely to be identified for improvement than those without these subgroups, resulting in some of the best schools in a State being identified as needing improvement.

Additional policy accommodation came in response to parts of the law that were not working well, and if strictly enforced would mean the loss of Title I funds to many States. For example, as the deadline for having all teachers highly qualified approached in 2005-06, it became clear that States would not reach the 100 per cent goal. In October 2005, Secretary Spellings announced a policy change that allowed States additional time to meet the highly qualified teacher requirements (Spellings

2005). With these changes, NCLB shifted from being one national policy applied uniformly on all jurisdictions to one dependent on what each State could negotiate with the federal administration. Following the tradition of ‘marble cake’ federalism in education, the federal government seems ready to address State concerns or risk further eroding political support for the law.

An example of intergovernmental accommodation in the urban context was Chicago’s success in gaining federal approval to provide tutoring programs for students in schools that failed the AYP. Under NCLB, districts that did not meet AYP, including most large urban districts, were prohibited from providing supplemental instructional services after school to their students. The U.S. Department of Education required that Chicago replaced its own services with outside vendors in January 2005. Mayor Daley stepped in and put his political capital behind the district CEO’s decision to continue the district services. In a series of private meetings between the Mayor and the US Secretary of Education, compromise was reached. In return for the district’s continuation of its supplemental services, the city agreed to reduce barriers for private vendors to provide tutorial services. When the compromise was formally announced by Secretary Spellings in early September in Chicago, Mayor Daley hailed the efforts as the ‘beginning of a new era of cooperation’ across levels of government in education (New York Times, 9 February 2005, p. A11). Similar waivers were subsequently granted in such cities as New York City and Boston. Clearly, intergovernmental negotiation is likely to be intense over the implementation of NCLB in complex urban systems.

Building a reliable data tracking system

State and Local agencies face a capacity gap in data management and tracking system. The Data Quality Campaign, a non-governmental organisation, focuses on the necessary elements in creating a ‘robust longitudinal data system’ where each State would gather data on the same student while they are in school. The student data are matched with other data files, such as teacher records and instructional support programs. According to the Data Quality Campaign (2008), a robust State-wide data system, features ten elements. These include a unique State-wide student identifier that matches individual student achievement and other information; a unique teacher identifier that links teachers to their students; and individual student data collected from public school through college.

Based on a 2009 survey on the ten key elements of a rigorous, longitudinal data tracking system on individual student performance, the Data Quality Campaign found that twelve States have instituted all ten elements, while 34 States have eight

or more (Data Quality Campaign 2010a). While all fifty States reported that their data systems have State-wide student identifier for tracking academic performance (including graduation and dropout), only 24 States adopted a State-wide teacher identifier that matched students to teachers. In these States with more robust systems, the Governor's office has played a critical role. In Delaware, two-term Democratic Governor Thomas Carper successfully pushed through a comprehensive education accountability plan between 1993 and 2000. A key feature of Governor Carper's reform was to link individual student's test scores to teachers. The 2000 reform plan enabled a professional standards board to use students' test achievement as the basis for 'at least 20 percent of the performance reviews given to teachers, administrators, and other instructional staff members' (Sack 2000). Governor Carper's successor, Governor Ruth Ann Minner, continued to insist on using student achievement to hold teacher accountable (Johnston 2005)

Seeing the need to utilise the data system, the Data Quality Campaign started surveying the States on their policies and practices to use data to improve student achievement in 2009. The first survey, published in January 2010, focused on ten actions that link State-wide data across P-20 and the workforce, expand data access to key stakeholders, and ensure professional capacity to use data for instructional practices (Data Quality Campaign 2010b). The survey found that only eight States were tracking individual students across P-20 and the workforce sectors, an assurance required for Phase 2 State fiscal stabilisation funds. While ten States reported strategies in sharing individual student progress data with educators, fewer than half of the States provided aggregated data reports to key stakeholders. To meet the federal performance-based expectations, States must improve their data utilisation to support student success.

Facing classroom realities

The NCLB accountability standards faced implementation realities in a multi-layered policy system. District administrators, school principals, and teachers were in the 'trenches'. Based on their clients' needs and the organisational reality, educators often bring forth a different view on what can be done. As part of a larger study on NCLB implementation, Sunderman and her collaborators illuminated implementation challenges at the district and school level during the first two years. In *NCLB Meets School Realities: Lessons from the Field*, Sunderman and her colleagues (2005, p. ix) argued that the NCLB had expanded federal involvement in education by 'reaching far more deeply into core local and state education operations.' Federal requirement on annual testing of core subjects in the elementary grades is seen as directly shaping curriculum and instruction. To approach NCLB from the classroom level, the research team conducted a teacher

survey with schools that were either struggling or improving in Fresno CA and Richmond VA during spring 2004. The survey showed that teachers generally recognised the value of focusing on student achievement generally had increased the amount of time allocated to test the core subjects due to NCLB. However, they were less certain on how failing schools would be sufficiently motivated by sanctions alone. An overwhelming percentage of the surveyed teachers saw the importance of collaborating with experienced administrators and seasoned teacher mentors. In other words, teachers saw the need to balance sanctions with professional development, curriculum support, and committed administrators.

3.5 Broadening of performance-based federalism in the Obama era

Focus on school turnarounds

The Obama Administration continues the push for more direct district intervention in persistently low performing schools. In his proposal to reauthorize the federal law in elementary and secondary education, Secretary Duncan argued for four strategies to ‘turnaround’ the nation’s lowest performing 5 per cent of the schools (or approximately 5000 schools). The federal government has committed \$5 billion during 2010-12 to support these efforts. The four strategies tighten the approaches that were established under NCLB, allowing for fewer district options. More specifically, the Duncan strategies include:

- *Turnaround school* under a new principal who can recruit at least half of the teachers from the outside
- *Transformation school* that strengthens professional support, teacher evaluation, and capacity building
- *Restart school* will reopen as either a charter school or under management by organizations outside of the district
- *School closure* that results in moving all the students to other higher-performing schools.

In making its first School Improvement Grants (SIG) to support school turnarounds, the Obama administration allocated \$3 billion to over 730 schools in 44 States in December 2010. Of these schools, an overwhelming number of them (71 per cent) had chosen the ‘transformation’ option while very few decided to use either ‘restart’ (5 per cent) or ‘school closure’ (3 per cent). The remaining 21 per cent opted for the ‘turnaround’ option where the principal and a majority of the teaching staff were

replaced (Klein 2011). Equally important, only 16.5 per cent of the students in all the SIG schools were white, as compared to 44 per cent African American and 34 per cent Hispanic.

Carrot and stick approach on new reform assurances

The Obama administration has strengthened the NCLB-like accountability system by making new federal investment in public education. In return, the federal government requires State and Local government to meet a set of new expectations on reforming public education. The *American Recovery and Reinvestment Act* of 2009 (ARRA) redefined the latest federal involvement in public education in several aspects.

First, ARRA created the ‘State Fiscal Stabilization Fund’ program to save over 300 000 teaching jobs in public schools during the time when many States and districts instituted fiscal retrenchment. Second, in return for federal stabilisation support, States are expected to meet federal expectations on education reform. Among the reform assurance areas are: 1) more equitable distribution of well-trained, well-qualified teachers to address students with greater needs; 2) ongoing monitoring of student progress with a data system that links pre-K to college and career development; 3) developing and implementing standards on college- and career-ready standards; and 4) taking effective actions to turnaround the persistently lowest-performing schools. Third, ARRA substantially expands federal funding in several categorical program areas, including education for the disadvantaged children (or Title I program), IDEA program for special education students, and financial assistance for eligible college students.

Equally important, ARRA invites States to submit their best ideas on system transformation and school innovation for the national competition for the Race to the Top program. Delaware and Tennessee were selected as the first two grantees of the first round of Race to the Top competition in April 2010. The second round of competition resulted in awarding ten States and Washington DC.

The winning applications submitted by Delaware and Tennessee share several features in their approach on transforming public education. First, teacher accountability is prominent. A system of teacher evaluation is established. Student achievement becomes a ‘cornerstone’ of this new assessment system, according to the Tennessee education commissioner. Delaware will use the annual evaluation results to remove teachers who are rated as ‘ineffective’ for consecutive years. Second, a system of support is developed to enhance professional capacity. Delaware plans to hire 35 data coaches to train teachers using data for instructional

improvement. The State also will hire 15 ‘development’ coaches to support principals and teachers in highest need schools. Third, external partners, such as Mass Insight Education and Research Institute, will be brought in. Fourth, the two States were successful in gaining approval from key stakeholders on the reform agenda in the long term. During the application, then Tennessee Governor Phil Bredesen was able to gain the endorsement on the application from all the gubernatorial candidates. To ensure institutional commitment, the two States set up administrative offices to oversee program implementation. Tennessee opened an ‘achievement school district’ office and Delaware set up a project management division to monitor the implementation of the reform initiatives. Finally, the two States show their support for expanding innovation. Tennessee recently passed a legislation that increases the number of charter schools and broadens student eligibility in school choice.

Federal–local alignment on innovation

To a large extent, ARRA goals align closely with State and Local priorities. Given the economic recession and substantial State budgetary cuts, State and Local governments strongly welcomed the Stabilisation program as a necessary means to fill the gap in teaching and other professional positions. U.S. Secretary of Education Arne Duncan pointed out that the Stabilisation program saved 330 000 teaching jobs across the country. These jobs would have been eliminated in the absence of the federal ARRA funds. The Congressional Budget Office estimated that the ARRA intergovernmental transfers have a multiplier effect as high as 1.9 times when federal dollars are used by State and Local agencies (Congressional Budget Office 2009).

Further, States have made efforts to take on innovative initiatives in recent years. Eighty per cent now have legislation supportive of the creation of charter schools. Charter expansion tends to promote the diverse provider reform model, which aims at system-wide shift to offer a broader mix of service providers as a strategy to raise student performance. Chicago, New York, and Philadelphia provide prominent examples of this approach (Wong and Wisnick 2007). As table 5 suggests, the growing charter school section involves diverse organisations. During 2009-10, there were 4903 charter schools enrolling about 1.67 million students. Of these schools, 492 are managed by Education Management Organizations (including for-profit and non-profit organizations) and 573 are managed by Charter Management Organizations. The diverse provider approach is likely to expand in the coming years.

With the widening use of corrective actions under NCLB, the issue of holding schools and teachers accountable has gained public support. In several States, governors and legislatures are beginning to experiment with differential compensation for educators. Florida and Texas, for example, provide individual cash bonuses to teachers for standardised test results. Arizona, Minnesota, and North Carolina connect part of the teacher salaries to student achievement. In Minneapolis and Denver, union leadership actively participated in negotiating with the management to redesign the teacher compensation package. Denver's ProComp Agreement did not eliminate collective bargaining. Instead, it gained voters' approval for new taxes to pay for the expanded salary schedule that takes into account four factors: knowledge and skills; professional evaluation; market incentives; and student growth. In response to State and Local interest in compensation reform, the federal government has expanded its investment in supporting alternative compensation initiatives. As shown in table 6, the federal Teacher Incentive Fund initiative has grown from \$99 million to \$443 million over the last 5 years. In other words, the policy conditions created by NCLB and ARRA will continue to facilitate State and Local innovation.

3.6 Can federal activism on accountability be sustained?

Federal activism on education accountability is now at a cross roads as President Obama pushes forward his own education agenda. On one hand, the Obama Administration has shown its strong support for performance-based accountability. Both the President and his Secretary of Education, for example, take the position that student performance matters in teaching employment and compensation. Recent public opinion polls indicate a clear majority supportive of teacher accountability. On the other hand, the sustainability of the new policy paradigm meets implementation challenges in the decentralised system. There are legal challenges filed against the annual testing requirements and other federal provisions. In Congress, federal education reform has shown partisan polarisation. The economic stimulus package, which included federal school aid, won Congressional passage without a single Republican vote in the House and only three Republican votes in the Senate. There remains political concern about federal infringement of State and Local control in education.

To students of implementation, the controversy, conflict, and resistance should come as no surprise. After all, when federal expectations are ambitious and encompassing, as they are in the federal NCLB, the organisational routine and the political status quo are being called into question. When lofty goals meet the

operational reality of federalism, we are likely to see implementation tension and intergovernmental conflict (Pressman and Wildavsky 1973; Peterson, Rabe and Wong 1986). At the same time, with the passage of time, implementation problems are likely to become more manageable as stakeholders at all levels of the government adjust their expectations (Peterson, Rabe and Wong 1986). As NCLB matures, the process of borrowing strength is likely to facilitate the institutionalisation of new policy norms and operational procedures. As NCLB ages, can the federal system reconcile the early conflicts? Does NCLB facilitate creative mechanisms to resolve these conflicts?

It remains to be seen how the performance-based paradigm will be fully institutionalized in our intergovernmental policy system. After all, categorical management remains highly routinised in the way government operates at all three levels. Nonetheless, given growing public concerns on school performance and the adoption of innovative practices, the new politics of accountability has elevated the federal role in education, an arena where States have always played a dominant role. Regardless of the future of the *No Child Left Behind Act* and the reform assurances established in the Obama administration, federal benchmarking in education has earned a prominent place in American federalism.

Table 3.1 Per pupil spending in public schools by sources of revenue in the US, 1996–2007

	<i>Per pupil expenditure</i>		<i>Source of revenue</i>		
	<i>Constant 2008 dollars</i>	<i>% increase over previous period</i>	<i>Federal %</i>	<i>State %</i>	<i>Local %</i>
1995-1996	8833	–	6.6	47.5	45.9
1998-1999	9535	7.9	7.1	48.7	44.2
2001-2002	10 443	9.5	7.9	49.2	42.9
2004-2005	9978	-4.5	9.2	46.9	44.0
2006-2007	11 941	19.7	8.5	47.6	43.9

Sources: US Department of Education, National Center for Education Statistics, *Biennial Survey of Education in the United States, 1919-20 through 1955-56*; *Statistics of State School Systems, 1957-58 through 1969-70*; *Revenues and Expenditures for Public Elementary and Secondary Education, 1970-71 through 1986-87*; and Common Core of Data (CCD), 'National Public Education Financial Survey', 1987-88 through 2006-07.

Table 3.2 US federal expenditures for elementary and secondary education

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Special-needs programs (in millions of 2005 dollars)										
Education for the disadvantage	8853.9	8763.1	9367.0	7839.4	9673.1	9541.2	10 039.3	11 944.1	12 909.3	14 638.2
Special education	4010.8	4022.2	4383.4	5209.7	5612.9	6416.0	7599.3	9012.1	10 079.7	10 226.5
Head start	4443.7	4843.6	5208.9	5460.4	5973.6	6841.0	7096.6	7076.2	7003.9	6843.2
Child nutrition programs	9802.3	10 099.6	10 262.1	10 407.3	10 835.6	10 528.0	11 131.7	11 493.3	11 586.1	12 163.9
Bilingual education	229.7	220.6	247.9	634.9	563.0	494.4	na	na	na	na
Native American education	96.3	68.1	63.1	76.0	74.1	93.2	112.8	123.1	118.28	129.9
<i>Subtotal</i>	27 436.7	28 017.2	29 532.4	29 627.7	32 732.3	33 913.8	35 979.7	39 648.8	41 697.3	44 001.7
Percent changes in special-needs programs over previous period	-	2.1	5.4	0.32	10.5	3.6	6.1	10.2	5.2	5.5
Federal spending for elementary and secondary education										
<i>Total (millions of 2005 dollars)</i>	43 818.4	43 171.5	44 914.5	46 818.0	49 685.8	53 842.0	57 270.0	62 914.4	64 775.9	67 959.2
Percent change over previous period	-	-1.5	4.0	4.2	6.1	8.4	6.4	9.9	3.0	4.9
Special-needs programs as percentage of federal spending (millions of 2005 dollars)	62.61	64.90	65.75	62.71	65.88	63.10	62.82	63.02	64.37	64.75

na not available.

Source: US Department of Education, National Center for Education Statistics, *Digest of Education Statistics*, 2005 (table 358), 2004 (table 366), www.nces.ed.gov.

Table 3.3 Meeting the AYP: state variation in starting points and AYP target patterns as defined in a sample of the 2002-03 accountability plans

State	AYP target patterns	Starting points		
		Grade	Language arts	Math
Arizona	Steady Stair-Step (Two Year Increments)	3	44%	32%
		3	32%	20%
		5	31%	7%
		High School	23%	10%
California	Equal Yearly Goals	2-8	13.6%	16.0%
		High School	11.2%	9.6%
Delaware	Equal Yearly Goals	All	53.9%	30.0%
Florida	Steady Stair-Step (Three Year Increments)	All	30.68%	37.54%
Kentucky	Equal Yearly Goals	Elementary	47.5%	22.73%
		Middle	45.6%	16.51%
		High School	19.26%	19.84%
Massachusetts	Steady Stair-Step (Two Year Increments)	All	39.7%	19.5%
Michigan	Steady Stair-Step (Three Year Increments) & Accelerating Curve from 2010-2014	Elementary	38%	47%
		Middle	31%	31%
		High School	42%	33%
New Hampshire	Steady Stair-Step (Three Year Increments)	3-8	60%	64%
		High School	70%	52%
New Jersey	Steady Stair-Step (Three Year Increments)	4	68%	53%
		8	58%	39%
		11	73%	55%
New Mexico	Equal Yearly Goals	All	37%	16%
New York	Steady Stair-Step (Three Year Increments)	4	123 AMO	136 AMO
		8	107 AMO	81 AMO
		11	142 AMO	132 AMO
North Carolina	Steady Stair-Step (Three Year Increments)	3-8	69%	75%
		10	52%	55%
North Dakota	Equal Yearly Goals	4	65.1%	45.7%
		8	61.4%	33.3%
		12	42.9%	24.1%
Ohio	Steady Stair-Step (Three Year Increments) & Equal Goals from 2010-2014	All	40%	40%

(continued next page)

Table 3.3 (continued)

State	AYP target patterns	Starting points		
		Grade	Language arts	Math
Oklahoma	Steady Stair-Step (Three Year Increments)	All	622 API, 1500 is 100%	648 API, 1500 is 100% proficient
Oregon	Steady Stair-Step (Three Year Increments)	All	40%	39%
Pennsylvania	Steady Stair-Step (Three Year Increments)	All	45%	35%
Texas	Equal Yearly Goals	All	46.8%	33.4%
Virginia	Steady Stair-Step (Two Year Increments)	All	60.7%	58.4%
Washington	Equal Yearly Goals	4	53.8%	30.3%
		7	30.8%	17.6%
		10	49.5%	25.4%

AMO Annual Measurable Objective. **API** Academic Performance Index.

Source: Selected 2003 Approved State Accountability Plans, <http://www.ed.gov/admins/lead/account/stateplans03/index.html>.

Table 3.4 Number of states by percentage of schools that met the adequate yearly progress (AYP)

% schools meeting AYP for each state	2002–03	2003–04	2004–05
0–25%	2	2	0
26–50%	5	1	6
51–74%	22	18	16
75–100%	13	25	27
na	9	5	2

N=51, which includes 50 states and Washington, DC. **na** Not available.

Source: Compiled from data accessed at website at each of the state education agencies.

Table 3.5 Number of charter schools managed by ‘diverse service providers’

<i>Provider Type</i>	<i>2007-08</i>	<i>2009-10</i>
Freestanding	3421	3838
Education Management Org	454	492
Charter Management Org	436	573
Total	4311	4903
Total Student Enrollment	1 289 449	1 665 779
Percent White	38.3	38.9
Percent Black	31.4	32.2
Percent Hispanic	24.2	23.2

Source: National Alliance for Public Charter Schools, various reports.

Table 3.6 Federal incentives in supporting performance-based compensation for teachers and principals

	<i>FY 2006</i>	<i>FY 2009</i>	<i>FY 2010</i>
Federal appropriation	\$99 million	\$200 million	\$442 million
No. new awards	33	20	62

Source: US Department of Education, Teacher Incentive Fund, <http://www2.ed.gov/programs/teacherincentive/awards.html>.

References

- Balfanz, R., Legters, N., West, T.C. and Weber, L.M. 2007, ‘Are NCLB’s Measures, Incentives, and Improvement Strategies the Right Ones for the Nation’s Low-Performing High Schools?’, *American Educational Research Journal*, vol. 44, no. 3, pp. 559–93.
- Betts, J. 2007, ‘California: Does the Golden State Deserve a Gold Star?’ in Hess, F.M. & Finn, C.E. (eds.) *No Remedy Left Behind: Lessons from a half-decade of NCLB*, AEI Press, Washington, DC.
- Center on Innovation and Improvement 2006, *Report on Restructuring*, Presentation at the Institute for School Improvement and Education Options, 25–26 September.
- Christman, J., Gold, E. and Herold, B.B. 2006, *Privatization ‘Philly Style’: what can be learned from Philadelphia’s diverse provider model of school management* (updated edition with important new information and findings)? Research for Action, Philadelphia.

-
- Congressional Budget Office 2009, 'Memo by Douglas W. Elmendorf, Director of Congressional Budget Office to Senator Charles Grassley on the Estimate of the Economic Effects of ARRA', 2 March.
- Congressional Research Services 2009, *Economic Stimulus: issues and process*. Report conducted by Jane G. Granelle, Thomas L. Hungerford, and Marc Labonte 3 April.
- Data Quality Campaign 2008, Information retrieved from http://www.dataqualitycampaign.org/survey_results/index.cfm?&print=Y on 15 April 2008.
- Data Quality Campaign 2010a, *DQC 2009–10 Survey Results Compendium — 10 Elements*, January.
- 2010b, *Inaugural Overview of States' Actions to Leverage Data to Improve Student Success*, January.
- DeBray, E. 2006, *Politics, Ideology and Education: Federal policy during the Clinton and Bush administrations*, Teachers College Press, New York.
- Eggen, D. and Glod, M. 2009, 'President Praises Results of "No Child" Law', *Washington Post*, 9 January, p. A02.
- Greene, J.P. and Winters, M.A. 2006, 'Getting Ahead by Staying Behind,' *Education Next*, vol. 6, no. 2.
- Hess, F. and Petrilli, M. 2006, *No Child Left Behind*, Peter Lang, New York.
- Hochschild, J. 2003, 'Rethinking Accountability Practices' in Paul E. Peterson and Martin R. West (eds) *No Child Left Behind? The politics and practice of school accountability*, Brookings Institution Press, Washington, DC.
- Janofsky, M. 2005, 'Judge Rejects Challenge to Bush Education Law,' *New York Times*, 24 November, p. A14.
- Johnston, R. 2005, 'Math Specialists Sought for Del. Middle Schools', *Education Week* 2 February.
- Kim, J.S. and Sunderman G.L. 2005, 'Measuring Academic Proficiency under the No Child Left Behind Act: implications for educational equity', *Educational Researcher*, vol. 34, no. 8, pp. 3–13.
- Klein, A. 2011, 'Turnaround-Program Data Seen as Promising though Preliminary', *Education Week*, 11 January. Retrieved from <http://www.edweek.org/ew/articles/2011/01/12/15turnaround-2.h30.html?qs=school+turnaround>.
- Lowi, T. 1964, 'American Business, Public Policy, Case Studies, and Political Theory', *World Politics*, vol. 16, pp. 677–715.

-
- Manna, P. 2006, *School's In: federalism and the national education agenda*, Georgetown University Press, Washington DC.
- 2010, *Collision Course: federal education policy meets State and Local realities*, Congressional Quarterly Press, Washington DC.
- McGuinn, P. 2005, 'The National Schoolmarm: No Child Left Behind and the new educational federalism', *Publius*, vol. 35, no. 1, pp. 41–68.
- 2006, *No Child Left Behind and the Transformation of Federal Education Policy, 1965–2005*, University of Kansas Press, Lawrence KS.
- Mintrop, H. 2004, *Schools on Probation: how accountability works (and doesn't work)*, Teachers College Press, New York.
- and Trujillo, T. 2005, 'Corrective Action in Low Performing Schools: lessons for NCLB implementation from first-generation accountability systems', *Education Policy Analysis*, vol. 13, no. 48.
- Oates, W. 1972, *Fiscal Federalism*, Harcourt Brace Jovanovich, New York.
- Peterson, P. 1995, *The Price of Federalism*, Brookings Institution Press, Washington DC.
- , Rabe, B. and Wong, K. 1986, *When Federalism Works*, Brookings Institution Press, Washington DC.
- Pressman, J. and Wildavsky, A. 1973, *Implementation: how great expectations in Washington are dashed in Oakland; or, why it's amazing that federal programs work at all; this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hope*, University of California Press, Berkeley CA.
- Public Agenda 2006, <http://www.publicagenda.org/issues>.
- Rossi, R. 2004, 'Sun-Times Insight: exploring the issues of the day,' *Chicago Sun-Times*, 8 July, p.16.
- Sack, J. (2000) 'Del. Ties School Job Reviews to Student Tests,' *Education Week* 3 May 2000.
- Spellings, M. 2005, *Letter to Chief State School Officers Regarding the 'Highly qualified teacher' Provisions of the No Child Left Behind Act*. Accessed November 2, 2005, from <http://www.ed.gov/policy/elsec/guid/secletter/051021.html>.
- Sunderman, G. 2006, *The Unraveling of No Child Left Behind: how negotiated changes transform the law*, The Civil Rights Project at Harvard University, Cambridge MA.

-
- and Orfield, G. 2006, ‘Domesticating a Revolution: No Child Left Behind and State administrative response’, *Harvard Educational Review*, vol.76, no. 4, pp. 526–556.
- Kim, J. & Orfield, G. 2005, *NCLB Meets School Realities: lessons from the field*, SAGE/Corwin Press, Thousand Oaks CA.
- US Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service 2007, *State and Local Implementation of the No Child Left Behind Act, Volume I—Title I School Choice, Supplemental Educational Services, and Student Achievement*, Washington DC.
- US Department of Education, National Center for Education Statistics, *Digest of Education Statistics*, 2004 and 2005, <http://www.nces.ed.gov>.
- US General Accountability Office 2009, *Recovery Act: States’ and Localities’ Current and Planned Uses of Funds While Facing Fiscal Distress*. Washington DC: US GAO, 09-829 July.
- 2010, *Recovery Act: One Year Later, States’ and Localities’ Uses of Funds and Opportunities to Strengthen Accountability*. Washington DC: U.S. GAO, 10-437 March.
- Walsh, M. 2010, ‘Court upholds dismissal of Conn.’s NCLB Suit’, *Education Week*, 14 July, Retrieved from http://blogs.edweek.org/edweek/school_law/2010/07/a_federal_appeals_court_has_4.html.
- Wong, K. 1990, *City Choices: education and housing*, SUNY Press, Albany NY.
- 1999, *Funding Public Schools: politics and policy*, University Press of Kansas, Lawrence KS.
- and Nicotera, A. 2006, *Successful Schools and Education Accountability*, Allyn and Bacon, New York.
- and Sunderman, G. 2007, ‘Education Accountability as a Presidential Priority: No Child Left Behind and the Bush presidency’, *Publius*, vol. 37, no. 3, pp. 333–50.
- and Wishnick, D. 2007, ‘Expanding the Possibilities: the diverse-provider model in large urban school districts’, in Robert Rothman (ed.), *City Schools*, Harvard Education Press, Cambridge MA.

4 Benchmarking health care in federal systems: the Canadian experience*

Patricia Baranek¹
University of Toronto

Jeremy Veillard²
Canadian Institute for Health Information

John Wright³
Canadian Institute for Health Information

4.1 Introduction

Benchmarking is broadly defined as the comparison of similar systems or organisations based on a recognised set of standard indicators (Wait and Nolte 2005). Distinctions are made between performance benchmarking and practice benchmarking; the former focuses on establishing performance standards while the latter is concerned with the underlying practices and search for best practices (Fenna, this volume). In the health sector, performance benchmarking is more prevalent, perhaps because health systems are complex and involve many institutions, sectors, payers and providers. The ongoing challenge has, therefore, been to link benchmarking to organisational change processes (Neely 2010).

Although international comparisons of health care systems date back to the 1930s, those early examples focused mainly on the structural characteristics of health care systems—such as the number of physicians and hospital utilisation data — or on a

* The authors would like to acknowledge the contributions of Kenneth Lam (University of Western Ontario), Indra Pulcins (Health Quality Ontario) and Adalsteinn Brown (University of Toronto).

¹ Patricia Baranek has an adjunct position in the Department of Health Policy, Management, and Evaluation at the University of Toronto.

² Jeremy Veillard is the Vice President, Research and Analysis at the Canadian Institute for Health Information.

³ John Wright is the President and Chief Executive Officer of the Canadian Institute for Health Information and former Deputy Minister of the Government of Saskatchewan.

few specific outcomes, such as average life expectancy at birth and maternal and child mortality. More recently, organisations such as the World Bank (1993), the World Health Organization (WHO 2000), the Organisation for Economic Co-operation and Development (OECD 2001) and the Commonwealth Fund (2007) have developed snapshots on cross-national health systems performance measurement. These reports have put international health system performance comparisons on the political agenda, raised awareness of performance issues, and resulted in initiatives to guide policies for the improvement of health care in individual countries (Veillard *et al.* 2010; Wait and Nolte 2005).

Consistent with international developments, there has been a significant increase in health system performance benchmarking in Canada over the last fifteen years. This is largely attributable to intergovernmental events of the late 1990s and early 2000s and the awakening needs of policy makers, health system managers, health care professionals and others to make informed comparisons so as to improve the safety, quality, timeliness and effectiveness of the health care system while ensuring Canadians are getting value for their tax dollars. More generally, this is an expression of the growing influence of management science and of the medical culture of evidence-based decision-making on health policy development and health system management (Pfeffer and Sutton 2006).

Initially, Canadian provinces and territories were reluctant to engage in performance benchmarking due to the fear of being compared, the perceived cost and design of data collection systems, and the aggressive timetables and workloads proposed. However, this lack of enthusiasm was overcome through political commitments to benchmarking that reflected pressure from the public, the media, and health care providers to get on with the job together with an infusion of incremental federal funding into the health system. The consequence has been a more accountable, transparent and informed system. Today, Canadian provinces and territories are committed to publishing more and better comparable data and extending the analysis from performance to practice benchmarking.

This study addresses the issue of health sector benchmarking in Canada so as to draw broader conclusions about the challenges and opportunities created by federal contexts. It first discusses the complexities of the Canadian federal system as it applies to the health sector. Second, it outlines examples of health sector benchmarking exercises conducted across the country by governments and authorities as well as other organizations. Three pan-Canadian benchmarking exercises are then explored in detail to highlight the characteristics (processes, outcomes, challenges and opportunities) of benchmarking experiences to date. The final section includes a discussion of lessons learned and potential future directions for benchmarking in the Canadian health sector.

The federal context in Canada

Canada is one of the world's largest decentralised federations, both in terms of geography and fiscal arrangements. It is a constitutional federation in which the division of power is enforced by the courts. The Constitution Act divides the responsibilities of government between the federal and provincial governments; the three territories are creations of the federal government. The federal government was granted unlimited taxing powers, while the provinces were limited to direct taxes within their jurisdiction. Provinces are highly protective of their constitutionally assigned jurisdictions.

Constitutional framework

With respect to health care, Section 92(7) of the Constitution Act gives the provinces exclusive jurisdiction over the 'establishment, maintenance, and management of hospitals, asylums, charities and eleemosynary institutions in and for the province, other than marine hospitals' and the federal government responsibility for marine hospitals and quarantine. The federal government also has jurisdiction over certain groups of individuals including Aboriginal peoples, veterans of the Canadian armed forces, the Royal Canadian Mounted Police, inmates of federal penitentiaries and refugee claimants. However, the different orders of government must coordinate their actions on health care as many issues cut across the jurisdictional boundaries originally defined under the Constitution Act (Simeon and Papillon 2006; McLean 2003).

A defining feature of Canadian health care is the Canada Health Act which was introduced by the federal government in 1984. It defines 'insured health services' to include hospital services, physician services and surgical-dental services provided to insured persons. Federal transfers for health care are conditional on the provinces and territories meeting the five principles or national standards of the Canada Health Act. Breach of these standards may result in a reduction or withholding of the federal cash contribution to the province in proportion to the gravity of the breach. The five principles of Canadian medicare are:

1. **Public administration:** a province's health plan must be administered on a not-for-profit basis by a public authority
2. **Comprehensiveness:** all medically necessary services rendered by a physician or surgeon must be covered
3. **Accessibility:** reasonable access to insured services by insured persons on uniform terms and conditions must be provided

-
4. Universality: medicare must be available to all provincial residents on equal terms and conditions
 5. Portability: benefits outside of the insured resident's province but within Canada must be made available.

The Canadian health care system is a collection of ten provincial, three territorial and one federal system with a core set of common programs and services as defined by the Canada Health Act. While it is mandatory to provide publicly funded medically-necessary hospital and physician services, jurisdictions also provide a broad array of other services such as home and long term care; mental health and addictions programs; and prescription drugs for specific groups. Private insurance and out-of-pocket payments cover items such as prescription drugs provided in the community and dental and vision care for populations other than the low-income or the elderly (CHSRF 2005).

The system is largely publicly funded (70 per cent public and 30 per cent private) covering medically necessary hospital care and physician services. In 2011, total health care spending in Canada (public and private) reached \$200.5 billion, more than 60 per cent more than a decade ago in real terms and approximately 11.6 per cent of 2011 GDP, of which 29.1 per cent was spent on hospitals, 14.0 per cent on physicians, 16.0 per cent on drugs, 6.3 per cent on public health and 10.0 per cent on other institutions. The remaining 24.6 per cent was spent on other professionals, administration, research and other health care goods and services. At the provincial level, some provinces are spending over 40 per cent of their operating budgets on health care and spending continues to rise faster than revenues (CIHI 2011).

The planning and delivery of health care is generally the responsibility of regional health authorities in seven provinces; local health integrated networks in Ontario (purchasing and planning of care only); and centralised systems in Alberta and Prince Edward Island. Each of the Territories has its own health region. Regional health authorities are devolved entities, created by the jurisdictions to increase local engagement in decision-making and to ensure that health care planning and service delivery are responsive to community needs.

Fiscal transfer programs for health care

Since 1919, the federal government has transferred funds to the provinces to finance portions of their health care systems and to ensure comparable standards of care. These transfers evolved from specific purpose cost-sharing grants in the 1930s and 1940s to broader cost-sharing mechanisms in the 1950s and 1960s and to a more

mature system of formula-based, per-capita, unconditional grants in the latter part of the 1970s.

Federal transfer payments have often created tension between the federal and provincial governments. Federal concerns focused on the perceived absence of provincial transparency and accountability and the risk that funds would be used for non-health related initiatives. Provincial concerns related to the possibility that unilateral decisions could be made by the federal government without consultation. Provinces and territories generally do not welcome federal intrusions unless they come with financial resources with few or no strings attached.

4.2 Health system renewal and the introduction of benchmarking

In the early 1990s most jurisdictions moved to reduce significantly or eliminate their fiscal deficits. Expenditure restraint initiatives throughout the country led to health care program restructuring. In 1995-96 the federal government reduced health care transfers in an effort to address its fiscal situation. By the late 1990s there was a marked improvement in the fiscal situation of most jurisdictions with many having balanced or surplus budgets. However, health care wait times had increased and the quality of care was perceived to have deteriorated, resulting in a national sense of urgency to improve the timeliness and quality of care.

In 2000 the Prime Minister and the provincial and territorial Premiers (collectively the First Ministers) reached agreement on a \$23.4 billion federally funded package of initiatives to strengthen and renew Canada's publicly funded health care services. Of particular note was a commitment to expand the sharing of information on best practices and to report regularly to Canadians on health status, health outcomes and the performance of publicly funded health services. Ministers of Health were charged with the responsibility to 'collaborate on the development of a comprehensive framework using jointly agreed comparable indicators such that each government will begin reporting by September, 2002.' Comparable indicators were to be developed in the following areas: health status; health outcomes; and quality of health care services (Health Canada 2000).

In late 2002, public reports were made available by each jurisdiction containing up to 67 indicators. However, not all indicators were directly comparable and data quality was suspect in many circumstances resulting in an inability to benchmark provincial and territorial health systems.

In 2003, the federal, provincial and territorial governments signed the ‘First Ministers’ Accord on HealthCare Renewal’. The Accord provided \$36.8 billion over five years to the provinces and territories to improve the accessibility, quality and sustainability of the public health care system and to enhance transparency and accountability. A new federal transfer mechanism (the Canada Health Transfer), along with the creation of the independent Health Council of Canada (with a mandate to monitor and make annual public reports on the implementation of the Accord), were among the initiatives. In addition, the Accord established parameters for an enhanced accountability initiative, beyond that of the 2000 agreement, focused on the development and reporting of comparable indicators around four themes: access (13 indicators), quality (nine indicators), sustainability (nine indicators) and health status and wellness (five indicators). These indicators were to be reviewed and approved by stakeholder groups and external experts so as to ensure their validity (Health Canada 2003).

In 2004 the First Ministers signed the ‘Ten-Year Plan to Strengthen Health Care’ wherein the federal government committed \$41.3 billion in additional funding, including targeted dollars for wait times reduction. The Plan also included an agreement to expand the number of comparable indicators and to develop evidence-based benchmarks for medically acceptable wait times for cancer and heart surgeries; diagnostic imaging procedures; joint replacements; and sight restoration surgery. All governments agreed to report to their residents on health system performance including the elements outlined in the Plan (Finance, Government of Canada; Simeon and Papillon 2006).

The 2004 agreement also specifically recognised an asymmetrical federalism that would allow for the existence of specific agreements for any province. In this instance, the agreement specifically recognised the distinct needs of Quebec. Quebec was to apply its own wait times’ reduction plan; issue its own report to Quebecers; and use federal funding to implement its own plans for renewing Quebec’s health care system (Health Canada 2004).

4.3 Intergovernmental coordination

Canada has developed a hierarchical structure of intergovernmental committees that usually include representation from all jurisdictions. These committees provide a forum for the constituent units of the federation to communicate; consult; harmonise their policies and programs; coordinate their activities; resolve conflict; and, in some instances, develop policy jointly. At the apex is the Conference of First Ministers. Within the health sector, the focal points are the Conference of Ministers of Health and the Conference of Deputy Ministers of Health which are supported by

committees of officials. The two conferences are each co-chaired by the federal government and a province (usually rotating annually). Provincial and territorial Ministers and Deputy Ministers traditionally meet prior to engaging their federal counterparts in an effort to coordinate agendas and develop a common front. The value of these pre-meetings has been questioned but is strongly supported by Quebec and Alberta. Generally, there is an ongoing degree of tension as between the federal government and the provinces and territories.

Ad hoc intergovernmental committees are often established to address specific issues. For example, to implement the comparable indicator reporting requirements of the Accords, a collaborative steering group of Deputy Ministers of Health was formed along with a working group. Once sufficient progress had been made in terms of organising, collecting, analysing and reporting of the indicators, the steering and working groups were disbanded and the ongoing responsibilities were devolved to two existing agencies, Statistics Canada and the Canadian Institute for Health Information (CIHI).

Interprovincial cooperation is fostered through a number of formal processes and organisations. For example, the Ministers and Deputy Ministers of the four Atlantic provinces meet regularly to discuss issues while their counterparts in the Western provinces and territories meet on an ad hoc basis. Additionally, officials often convene collectively or bilaterally to address specific items such as health human resource issues and pharmaceutical purchasing arrangements. Governments, with the exception of Quebec, also jointly fund organisations such as the Canadian Agency for Drugs and Technologies in Health (CADTH — health technology assessments) and the Canadian Blood Services (CBS — collection and distribution of blood products).

Quebec is highly selective in its participation in intergovernmental forums, initiatives and organisations. Although Quebec officials attend most intergovernmental meetings, their contributions to the dialogue are selective. Equally, although Quebec chooses not to participate in many pan-Canadian initiatives and organisations, it is a data contributor and participant in benchmarking exercises. Quebec's isolationist approach has been an ongoing frustration for some jurisdictions that would prefer to see an active, pan-Canadian role for Quebec in developing common policy positions to take to the federal government and solutions to issues; the insight and expertise that Quebec officials can bring to the table are felt lost and economies of scale that might be obtained forgone.

The degree of intergovernmental cooperation is often driven not only by issues and politics but also by the personalities of officials and Ministers and fiscal circumstances. As governments, Ministers and Deputy Ministers of Health change

overtime, the degree of cooperation can change as well; intergovernmental relations can be quite fluid.

4.4 An overview of health system benchmarking initiatives in Canada

Canada has made considerable strides in the development of comparable health care indicators at the institutional, regional, provincial, and pan-Canadian levels. This has led to a reasonably comprehensive array of performance based benchmarking initiatives, particularly within the acute care sector. However, practice benchmarking based on best practice or medical evidence is relatively new.

At the pan-Canadian level, health data collection, monitoring and reporting have moved from federal ministries and agencies to the Canadian Institute for Health Information (CIHI), an independent institution which, through collaborative processes with all governments, is the developer of quality performance indicators and the custodian and reporter of the data.

Early days of data collection

Since 1963 Health Canada has held the National Health Accounts and provided some data on health systems. Expenditures were initially compiled only for personal health care — including hospitals, prescribed drugs, physicians, dentists and other professionals. Data were also gathered from a number of sources, including an annual hospital survey (Statistics Canada); a retail drugstore survey on prescription drugs (jointly by Statistics Canada and the Canadian Pharmaceutical Association); and income tax statistics to estimate income of private-practice physicians, dentists and other professionals. Eventually nursing homes, non-prescription drugs, health appliances and other health expenditures (public health, capital expenditures, administration of insurance programs and research) were added to personal health care in the National Health Accounts.

Canadian Institute for Health Information

Recognising the importance of monitoring the performance of the health system across the country through the use of standardised indicators, the Canadian Ministers of Health established the Canadian Institute for Health Information in 1994. Its mandate is ‘to serve as the national mechanism to coordinate the development and maintenance of a comprehensive and integrated health

information system in Canada’ and to be a ‘source of unbiased, credible and comparable health information’. Funded through five-year agreements with Health Canada (80 per cent of total funding) and bilateral agreements with the provinces (including Quebec) and territories (17 per cent of total funding), the Institute is an independent, not-for-profit organisation. Although it gives precedence to the priorities of all governments, CIHI determines to a large extent its own priorities. At the same time, it consults with jurisdictions to ensure their support and efforts in providing high quality data.

Through data provided by hospitals, regions, medical practitioners and governments, CIHI tracks activity and performance in many areas. Its annual and ad hoc reports cover health care services, health spending, health human resources and population health. It is the primary source of pan-Canadian health care indicators that are used for performance benchmarking analysis.

CIHI works with stakeholders in developing and promoting standardised, comparable indicators and reports. For example, in 2006 Canada Health Infoway, whose mandate is the development and acceleration of the use of electronic health records across Canada, and CIHI launched a pan-Canadian coordination function to support and sustain health information standards on a national scale. The collaborative has generated 20 standard-development projects that are either completed or underway (Canada Health Infoway).

The Canadian Health Information Roadmap Initiative

The Canadian Health Information Roadmap Initiative, first launched in 1999 and renewed in subsequent years, was a collaboration between CIHI, Statistics Canada, Health Canada and other stakeholder groups at the national, regional and local levels. The initiative was federally funded following recommendations from the National Forum for Health in 1997 and was to assist later in implementing the work earmarked by the First Ministers’ Accords discussed above. It was to develop performance indicators that answered two fundamental questions: how healthy are Canadians; and how healthy is the Canadian health system? The Roadmap Initiative consisted of several projects dealing with reports and indicators; integrated health services; health resources management; info-structure and technical standards; and population health (CIHI 2004).

Part of the Roadmap Initiative was the Health Indicators Project. One of the products from the Health Indicators Project is the Health Indicators Framework, which is based on a population health model with four dimensions: health status; non-medical determinants of health; health system performance; and community

and health system characteristics. Since 1998 the Roadmap Initiative, participants have collaborated to develop and implement the Framework (Arah et al. 2003). The Framework was recently endorsed by the International Organization for Standardization (ISO) as an international standard for performance measurement in the health sector (figure 4.1).

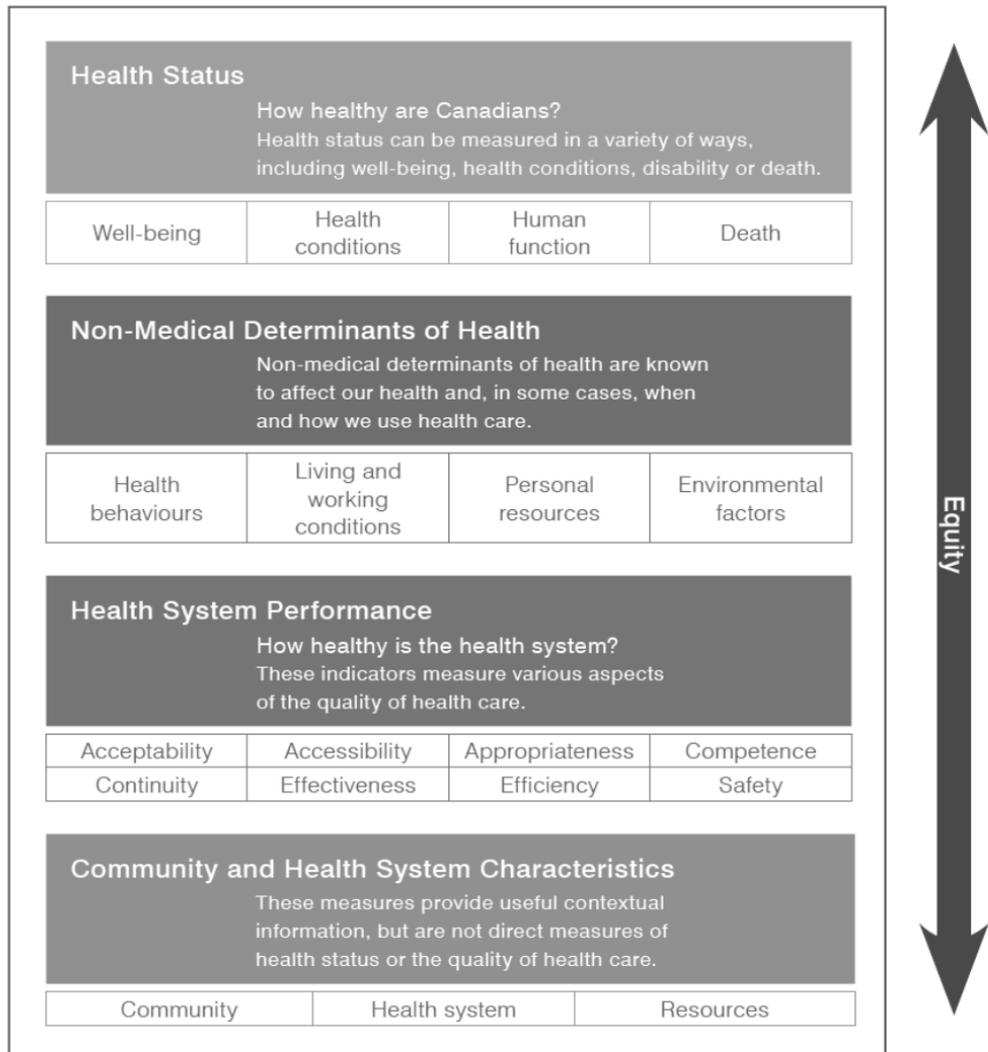
Other monitoring bodies

Although CIHI is the major body reporting on pan-Canadian health system performance, there are a number of other bodies that also monitor and report on the performance of the health system. At the pan-Canadian level, the Health Council of Canada reports on progress in improving the quality, effectiveness and sustainability of the health care system to all Canadians. Many provinces also have independent health quality councils to report to their constituencies on health system performance. Further, at the provincial level, there are numerous efforts to improve care and measure performance for specific conditions, such as the Alberta Cardiac Access Collaborative, the Cancer Quality Council of Ontario, the Ontario Cardiac Care Network and the Saskatchewan Chronic Disease Management Collaborative.

There are a number of not-for-profit, independent public policy centres, institutes, projects and think tanks that produce reports on Canadian health care performance, such as the Conference Board of Canada, the Institute for Clinical Evaluative Sciences, the Frontier Centre for Public Policy, the Hospital Report Research Collaborative, the Institute for Research on Public Policy, the Fraser Institute and the Canadian Centre for Policy Alternatives. And, academic researchers using CIHI data and other information also conduct studies comparing Canadian health system performance.

Figure 4.1 Health indicators: a framework

Health Indicator Framework



Sources: Canadian Institute for Health Information; Statistics Canada.

4.5 Illustrations of benchmarking activities in Canada

By exploring the goals, processes, outcomes, and challenges associated with three pan-Canadian initiatives, the evolution from a performance benchmarking approach (Health Indicators Project) to practices more aligned with practice benchmarking (Canadian Hospital Reporting Project and Hospital Standardized Mortality Ratios Reports) becomes clear.

The Health Indicators Project

As previously noted, the Health Indicators Project is a collaborative effort with a goal to provide reliable and comparable data on the health of Canadians, the health care system and the determinants of health. Over the period of the collaboration, the partners have held three National Consensus Conferences to develop a performance indicator framework and to determine a core set of indicators relevant to established health goals and strategic directions. These indicators are based on agreed upon benchmarks, guidelines and standards, collected using standardised data definitions and elements and available electronically across Canada to a national, provincial, regional or local level.

An intergovernmental advisory group was established to guide the project. Regional reference groups were created to provide expert advice on regional information needs, to ensure the quality and consistency of the indicator data and to provide guidance on the future development of the initiative. In addition, to extend the project reach and access to data, the Health Indicators e-publication was created.

Clinical data are obtained from data bases provided by all jurisdictions; from the Canadian Community Health Survey, which provides data at postal code levels; and from the Canadian Census. CIHI and Statistics Canada monitor these data; ensure data meet nationally agreed upon standards; and analyse the data. An annual report is released to policymakers, health system managers, researchers and the general public. Initially, the report was not made public to allow time for facilities and jurisdictions to validate their data and for CIHI to provide assistance to the regions on data interpretation and use. Everyone involved is expected and encouraged to explore and interpret their data in the context of local conditions, processes and experiences. Jurisdictions may also release their institutional, regional and provincial level-specific data on their websites.

The project depended on the collaboration and cooperation of governments, regional and local health organisations, key data custodians, and Canada's health research community. The incentive for participation was the support provided to

jurisdictions and authorities through the provision of reliable, standardised data with which to monitor, improve and maintain the health of their populations and the functioning of their health systems.

The comparable indicators are used to inform health policy, manage the health care system, improve the understanding of the factors that influence health and identify gaps in health status and outcomes for specific populations. Some provinces have incorporated the data into accountability agreements with their regional health authorities. In addition, the data are utilised extensively by provincial health quality councils and others to highlight top and bottom performing institutions within their jurisdiction.

Regional health authorities employ the data to gain a better understanding of their operations and how they compare with other authorities. The information has been used, for example, to accelerate change by comparing acute-care length of stay and wait times for surgical procedures. It has encouraged improved efficiency and effectiveness of operations by drawing attention to underperforming areas and helped create awareness of substandard care. It has resulted in quality and patient safety improvement projects; assisted in establishing workload productivity targets; and helped to avoid policy shifts when the data did not support it. Of particular note, the reporting of benchmarking data through the media has significantly enhanced public awareness of the relative quality of local health care services — which in turn has led to calls for improvements.

For example, for three years in a row, data from this project showed that hip fracture patients in the Winnipeg Regional Health Authority were waiting longer for surgery than in most other Canadian health regions. As a consequence, the health authority initiated a work plan to reduce wait times. It held continuing education sessions for staff; changed its practice of easing patients off blood thinners prior to surgery; reorganised surgery slates; and implemented a real-time information system to provide information about hip fracture patients waiting at every facility across the region. The result has been shorter wait times and better patient care.

The provision of benchmarking indicators alone does not ensure the integration of these tools into ongoing policy, planning and operational activities. CIHI provides extensive education workshops, technical information and reporting tools for managers and analysts. In the future, CIHI will introduce better business intelligence tools (e-reporting) to enable jurisdictions to incorporate this information into their decision-making processes more effectively. CIHI will also continue to expand and revise the set of indicators to reflect the changing needs of jurisdictions.

The Canadian Hospital Reporting Project (CHRP)

The Canadian Hospital Reporting Project (CHRP) is one of CIHI's current strategic initiatives and is based on the work initiated by Baker and Pink in the 1990s (Baker and Pink 1995). It is a project initiated in 2009 to look at facility-level comparative performance in acute-care hospitals. Its goal is to provide information that hospitals can use to identify areas of needed improvement, allowing hospitals to compare themselves with other institutions on clinical effectiveness, including outcomes and patient safety, and financial performance. All provinces and territories voluntarily agreed to participate and all Canadian acute care hospitals (over 600) are participating. Since April 2012, the results for 30 indicators (21 clinical indicators and 9 financial performance indicators) are publicly available at www.cihi.ca.

Participating jurisdictions completed a survey to determine their needs. Indicators were chosen through a rigorous selection process with input and agreement from expert groups consisting of researchers; policy makers; administrators; representatives from provider associations; and other stakeholders. Ten to fifteen indicators were chosen for each of the two dimensions and a number of different clinical and financial databases enable CIHI to populate the indicators.

Results are provided through an interactive web-based tool. Based on hospital profile information, CIHI has created peer groupings whereby each hospital is assigned to one of four standard peer groups. Based on hospital capacity, patient complexity, operations and resources, hospitals may also create their own custom comparator group. As a result, hospitals can compare their results to their regional, provincial, and national peers. In the future, the project will be expanded to include other dimensions of performance, such as patient experience, system integration and change and other health care sectors will be covered such as rehabilitation, mental health, long term care and continuing care.

Hospital standardized mortality ratios (HSMRs) reports

The publication of 'hospital standardized mortality ratios' (HSMRs) since 2007 by CIHI is associated with performance improvements at the facility level and adjustments to government policies and legislation requiring the indicator to be reported publicly. The HSMR is a summary measure adapted in Canada from the work of Sir Brian Jarman in the United Kingdom (Jarman *et al.* 1999). It is a ratio of the actual number of deaths in a hospital compared to the average Canadian experience, after adjusting for factors such as age, sex, diagnoses and admission status of patients. It provides hospitals with a starting point to assess mortality rates and to identify areas for improvement to reduce hospital deaths from adverse

events. Although not all deaths are avoidable, this indicator provides useful information in cases where they are avoidable. When tracked over time, the ratio can be a motivator for change, by indicating how successful hospitals or health regions have been in reducing inpatient deaths.

The use of these data for improving the quality of hospital care is illustrated by the examination of a septicaemia incidence by a health centre in Ontario. Their assessment based on HSMR data confirmed a delay in identifying sepsis, as well as inconsistencies in practices. As a result, it developed best practices and standardised orders for use in wards and emergency rooms. A rapid response team was also introduced for early recognition and treatment of cases. The result has been a continuing decline of septicaemia mortality rates for the hospital.

The activities of this facility as well as those of other hospitals across Canada followed the publication of HSMR data. Because further scrutiny showed that sepsis was a major cause of potentially preventable deaths, CIHI undertook a more detailed analysis to demonstrate how the HSMR data could be used for monitoring and quality improvement in Canadian acute care facilities (CIHI 2009). Many provinces have now adopted the indicator in their accountability agreements with facilities or regional health authorities and results are available now for all facilities including Quebec.

4.6 Discussion and lessons learned

Implementing meaningful benchmarking activities in the Canadian health system is complicated by the difficulty of comparing different health systems in a context of asymmetrical and at times strained relationships between orders of government. Nevertheless, this past decade has seen a significant expansion in the development, measurement and reporting of standardised indicators furthering performance benchmarking at all levels.

The size, complexity and cost of the tasks necessary to implement comparable indicator reporting as envisioned in the First Ministers' Accords was underestimated. Establishing data collection standards and methods, developing quality, comparative indicators and making the information broadly available have all required significant effort beyond that originally anticipated. Although there has been an ebb and flow over the last decade in the commitment to benchmarking reflecting changing health care priorities, all governments currently are committed to the process and regional health authorities even more so.

Although performance benchmarking continues to serve a practical function, the linking of performance data to quality improvement in Canadian health care systems, because of jurisdictional powers, is largely left to the constituent units. As a result, improvement in health systems is dependent on the political and social contexts of jurisdictions and the desires, skills and priorities of management. Still, there is interest in moving towards practice benchmarking where facilities and jurisdictions can compare their performance with those of their peers and extract and apply policy lessons to their own systems and projects. One such example is a collaboration of Canadian academic health science centers, CIHI and others to establish quality and patient safety practice benchmarking in their acute-care institutions. A key component of this exercise is for the participants to share and learn from best practices in each facility. Similarly, provincial health quality councils have encouraged regional health authorities to learn from each other through the sharing of best practices.

Challenges related to benchmarking in the health sector in federal systems

A number of challenges unique to the Canadian federal context will probably persist and place a limit on future health system benchmarking activities.

For example, the systems in each of the 14 jurisdictions are constantly evolving due to political agendas and efforts to control costs while improving quality, safety and access. Additionally, the past two decades have seen dramatic changes in the organisation and planning of health services, most notably in the creation and modification of regional bodies. These reforms have created difficulties in data aggregation and comparisons over time. To add to the complexity, these health care system changes are often on different time trajectories.

Privacy of health information has taken a key role on the policy stage and results in varying and sometimes unconnected pieces of legislation. Regional variations in determinants of health such as unemployment, education and poverty as well as different economic capacities can result in challenging comparisons across jurisdictions on health system functioning and outcomes. To mitigate this challenge, CIHI provides contextual information and support to jurisdictions to help them interpret their data. It also creates clear and accurate messages based on its findings for the media and the general public.

Data quality is an ongoing issue due to human error in coding; changes in coding practices; the lack of comparability of data sources; and issues related to the

submission of data. Accordingly, CIHI devotes significant resources to data quality reviews and improvements.

The gaming of data is always a potential challenge. Upcoming shifts towards activity based funding initiatives in several provinces suggest that gaming could become an even greater concern. In addition, data are not always available for meaningful, relevant indicator development for some sectors of health care and/or for all jurisdictions. Although data and benchmarking exercises are largely focused on acute care presently, considerable efforts are ongoing for the development of indicators and databases in other care sectors.

Due to the delay in receiving data from jurisdictions, the timeliness of reports is a challenge with results based on data that are generally one year old or more (with the exception of HSMR results and Emergency Rooms indicators for which monthly and quarterly results are available). CIHI is currently moving towards within-year data availability to allow facilities and authorities to review their results at any time despite the fact that such data may not as yet have gone through the full cycle of quality checks.

Finally, linking information to improvement requires careful consideration. Mis-use in the adoption of best practices from other jurisdictions has been well documented and include the selection of information to further political goals; the importation of modes or practices without validation; and differing and potentially contradictory motivations (Klein 1997).

The way ahead for benchmarking in the health sector in Canada

Developments in Canadian health care benchmarking are paving the way for future benchmarking practices. Overall, benchmarking in the health sector is expanding from performance benchmarking to practice benchmarking through the comparison of performance with peer groups and the learning from better performers.

The selection of benchmarks is becoming more focused and is increasingly driven by health systems' priorities and performance expectations. In line with international experience, performance measurement has become one basis for policy discussions concerning ways to improve health system performance (Veillard *et al.* 2010). From this perspective, a well-designed benchmarking system has the potential to guide policy development and can be used both prospectively and retrospectively. It can be used retrospectively and prospectively — supporting better understanding of past performance and the rationale behind certain performance patterns and helping to revise strategies for improving future performance (Nolte *et al.* 2006).

A focus on performance improvement and the close linkage of performance measurement and strategy should guide future benchmarking systems. For meaningful change, these systems should have certain characteristics.

- A strategic focus linking the design of the benchmarking initiatives with health system strategies ensures that policy lessons will be drawn in a way conducive to performance improvement.
- Data standardisation efforts are required to facilitate credible comparisons at both the Canadian and international levels.
- A policy focus rather than research focus implies that benchmarking systems should be driven by policymakers and system managers supported by experts and researchers.
- Translating performance information in easy ways for policy makers and managers to comprehend the information is important.
- Finally, sensitivity to political and contextual issues implies that interpretation of indicator data should not lose sight of the policy context within which they are measured, of the players involved in formulating and implementing policy, of the time lag needed to assess the impact of different policies and of aspects of health care that remain unmeasured by available data (Veillard *et al.* 2010).

Pressure to constrain public health care spending and the necessity to allocate resources in a way that promotes better health and economic growth are increasingly pushing Canadian jurisdictions to make better use of high quality data to compare performance and learn from one another. Despite ongoing concerns about data collection costs and complications related to the federal context, there is willingness by Canadian jurisdictions to collaborate towards health system performance improvement and better health. In other words, the fear of comparison has given way to the need for improvement. This evolution will require further investments in health information, better bridging research and analysis using the data with reviews of the available scientific literature on options for performance improvement and careful consideration of the choice and modalities of the benchmarks. Perhaps most importantly, it will require a shift in the culture of health care managers at every level of the system to one where they value and manage more extensively by data.

References

Arah, O., Klazinga, N., Delnij, D., Abroek, A., and Custers, T. 2003, 'Conceptual frameworks for health systems performance: a quest for effectiveness, quality,

-
- and improvement’, *International Journal for Quality in Health Care*, vol. 15, no. 5, pp. 377–98.
- Baker, R., and Pink, G. 1995, ‘A balanced scorecard for Canadian hospitals’, *Healthcare Management Forum*, vol. 8, no. 4, pp. 7–13.
- , Brooks, N., Anderson, G., Brown, A., McKillop, I., Murray, M. and Pink, G. 1999, ‘Health Care Performance Measurement in Canada: who’s doing what?’, *Healthcare Quarterly*, vol. 2, no. 2, pp. 22–26.
- Boadway, R. and Watts, R. 2004, *Fiscal Federalism in Canada, the USA and Germany*. Working Paper, no. 6, Institute of Intergovernmental Relations, Queen’s University at Kingston.
- Canada Health Infoway, ‘Health Information solutions are helping to transform health care in Canada,’ http://www.infoway-inforoute.ca/flash/lang-en/scguide/docs/StandardsCatalogue_en.pdf (accessed 22 May 2012).
- Canadian Health Services and Research Foundation 2005, ‘Myth: Canada has a communist style health care system’, http://www.chsrf.ca/mythbusters/html/myth18_e.php.
- Canadian Institute for Health Information 2009, ‘A Focus: a national look at sepsis’. http://secure.cihi.ca/cihiweb/products/HSMR_Sepsis2009_e.pdf.
- 2009, ‘Health Care in Canada 2009: a decade in review’, http://secure.cihi.ca/cihiweb/products/HCIC_2009_Web_e.pdf.
- 2009 Hospital Standard Mortality Rates, http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=hsmr_results_home_e.
- 2004, Annual Report 2003-2004, <http://dsp-psd.pwgsc.gc.ca/Collection/H115-2004E.pdf>.
- nd, *Canadian Hospital Reporting Project*, http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=indicators_chrp_e.
- nd, *CIHI — Taking Health Information Further*, http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=profile_e.
- nd, *Health Indicators*, http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=indicators_e.
- Finance, Government of Canada, *A History of the Health and Social Transfers*, <http://www.fin.gc.ca/fedprov/his-eng.asp>.
- Government of Canada 1999, *A Framework to Improve the Social Union for Canadians: an agreement between the government of Canada and the governments of the provinces and territories*.

-
- Health, Government of Canada 2004, *First Ministers' Meeting on the Future of Health Care, A 10-Year Plan to Strengthen Health Care*, http://www.hc-sc.gc.ca/hcs-sss/delivery_prestation/fptcollab/2004-fmm-rpm/index-eng.php.
- Jarman, B., Gault, S., Alves, B., Hider, A., Dolan, S., Cook, A., Hruwitz, B. and Iezzoni, L. 1999, 'Explaining Differences in English Hospital Death Rates Using Routinely Collected Data' *British Medical Journal*, vol. 318, pp. 1515–20.
- Justice, Government of Canada. *The Canada Health Act*, <http://laws.justice.gc.ca/eng/C-6/page-1.html>.
- Marshall, M.N., Shekelle, P.G. and Brook, R.H. 2000, 'The Public Release of Performance Data: what do we expect to gain? A review of the evidence', *Journal of the American Medical Association*, vol. 283, no. 14, pp. 1866–74.
- Klein, R. 1997, 'Learning from Others: shall the last be the first?', *Journal of Health Politics, Policy and Law*, vol. 22, no. 5, pp. 1267–78.
- Lockhart, W. 2010, Health Regions Report *Cards and Rankings*, <http://www.uregina.ca/admin/faculty/Lockhart/download.html>
- McLean, I. 2003, *Fiscal Federalism in Canada*. Nuffield College Politics Working Paper 2003-W17, University of Oxford.
- Ministry of Health and Long Term Care 2010, *The Excellent Care for All Act, 2010*, http://www.health.gov.on.ca/en/legislation/excellent_care/default.aspx.
- Neely, A. 2010, *Benchmarking: lessons and implications for health systems*. Unpublished report for the World Health Organization, European Region.
- Ogilvie, L., *SHA Implementation*, Canadian Institute for Health Information. www.oecd.org/dataoecd/2/27/20302730.pp.
- OECD 2001, *Health at a Glance 2001*, Organisation for Economic Co-operation and Development, Paris.
- 2010, *OECD Health Data 2010*, Organisation for Economic Co-operation and Development, Paris, www.oecd.org/health/healthdata.
- 2010, *OECD Health Data 2010: how does Canada compare*, Organisation for Economic Co-operation and Development, <http://www.oecd.org/dataoecd/46/33/38979719.pdf>.
- Pfeffer J. and Sutton, R. 2006, *Evidence-Based Management*, Harvard Business Review, Harvard, pp. 63–74.
- Simeon, R. 2002, *Federalism in Canada, a visitor's guide*, unpublished document.
- and Papillon, M. 2006, 'The Division of Powers in the Canadian Federation', in Akhtar Majeed, Ronald L. Watts, and Douglas M. Brown (eds), *Distribution*

of Powers and Responsibilities in Federal Countries, McGill–Queen’s University Press, Montreal and Kingston , pp. 91–122.

Veillard, J., Garcia-Armesto, S., Kadandale, S. and Klazinga, N. 2010, *International Health System Comparison: from measurement challenge to management tool*, in *Performance Measurement for Health System Improvement*, Cambridge University Press, Cambridge.

Wait, S. and Nolte, E. 2005, ‘Benchmarking health systems: trends, conceptual issues and future perspectives’, *Benchmarking: an international journal*, vol. 12, no. 5, pp. 436–48.

——— 2004, *Benchmarking: a policy analysis*, The Nuffield Trust, London.

WHO 2000, *The World Health Report 2000 — Health Systems: improving performance*, World Health Organization, Geneva.

5 Towards benchmarking in Germany

Gottfried Konzendorf¹
Federal Ministry of the Interior

Rabea Hathaway²
Federal Ministry of the Interior

A constitutional amendment effective 1 August 2009 gave constitutional status to comparative studies between different levels of public administration in Germany. The present article discusses this constitutional amendment and its implementation in the Federal and *Länder* administrations in Germany, where, to-date, only a handful of benchmarking experiments have occurred above the local level.

5.1 The basic structure of German federalism

The federal state

The fundamental principles of Germany's state structure are enshrined in the constitution of 1949. Among other things, the constitution states that the Federal Republic is a federal state (Article 20). Under Article 79(3) this principle cannot be changed even by a constitutional amendment.

The Federal Republic of Germany is a two-tier federated state. State authority is divided between the Federation and its constituent units, the 16 *Länder*. It should be noted that the *Länder* are neither provinces nor *départements* but States with their own executive authority, constitutions, parliaments and administrative structures. Thus the individual *Länder* act independently. The Federation has no authority to tell them how to carry out their tasks; it may exert direct influence only if it delegated the task to the *Länder*.

¹ Gottfried Konzendorf is the former Deputy Head of Division of the Division for Improved Legislation and Reduction of Bureaucracy.

² Rabea Hathaway is the former Deputy Head of Division: Modernisation of Administration and Shared Service Centres.

The *Länder* form the basis of authority in the political system. When responsibilities are to be divided between the Federation and the *Länder*, the Constitution assumes that the *Länder* are responsible unless otherwise specified. The Federation has responsibility for legislation, administration and jurisdiction only if expressly mentioned or implied in the constitution (Article 30). However, the *Länder* must abide by federal legislation, on which they have input via the *Bundesrat* (Federal Council), where their delegates sit as the second chamber of the federal parliament (see Fenna, this volume).

Legislative powers

Regarding the division of responsibility for legislation, the constitution provides for three types of legislative powers:

- exclusive legislative power at Federal level
- concurrent federal and *Länder* legislation
- exclusive legislative power at *Länder* level.

Exclusive legislative power at the federal level covers all tasks explicitly assigned to the federal level by Article 73 (together with Article 71) of the constitution or where implicit responsibility derives from the type of task or in connection with a directly assigned federal responsibility. These tasks include, for example, foreign affairs, defence, citizenship, currency, customs, foreign trade and payments, border protection, railways, aviation, postal and telecommunications services, copyright, counter-terrorism and nuclear energy.

Concurrent legislation includes the areas listed in Article 74 (together with Article 72) of the Constitution. They include civil law, criminal law, public welfare, economic law, labour law, food law and cartel law, social insurance, medicines, transport, environmental protection, university admission and degrees, and regional planning. If no federal law has been passed in these areas, the *Länder* are responsible. It should be noted that in certain areas, federal law is permitted only when necessary to establish equivalent living standards throughout Germany, or to maintain legal or economic unity in the national interest.

The exclusive legislative power of the *Länder*, finally, includes all areas not exclusively assigned to the federal level or subject to concurrent legislative power. This mainly affects schools and higher education, press and broadcasting law, and law on local authorities, police and regional planning.

Over the years, the Federation has adopted laws not only in areas where it has exclusive legislative power but also in most areas of concurrent legislative power.

The *Länder* are no longer authorised to make laws in these areas. However, this does not affect their other constitutionally guaranteed responsibilities. And via the *Bundesrat*, the *Länder* have a key role in the making of federal laws.

In addition to separate responsibilities for certain areas, there are tasks which can be carried out by the federal and *Länder* levels together —Article 91(a) and 91(b) of the Constitution). These shared tasks include the improvement of regional economic structures, agrarian structures and coastal preservation, and the promotion of research. The federation bears between 50 per cent and 90 per cent of the costs.

Administrative responsibility

Implementing the law also depends on the constitutional division of powers. If the *Länder* act in their own right, they are responsible for implementation. The *Länder* are also responsible for implementation when they are tasked with executing federal law (Article 84 of the Constitution) or on federal commission (Article 85). However, in these areas the federation may adopt laws (in some cases not binding) governing the authorities' organisation or administrative procedure, or adopt general administrative regulations, with *Bundesrat* consent, which are binding for the *Länder*. For example, general administrative regulations were adopted for authorisation procedures under environmental law; for issues related to enforcing the German Road Traffic Regulations; for guidelines on criminal proceedings and proceedings for the collection of fines; and for detailed provisions on the implementation of tax law.

Although all powers not specifically assigned to the Federation are reserved to the *Länder*, most legislation is passed at the Federal level. Laws are then usually implemented by the *Länder*. To important extent, then, German federalism embodies a different approach to dividing responsibilities from the American model (see Fenna, this volume).

In terms of state structure, municipalities (including 12 200 cities and communities, and 301 rural districts) are part of the *Länder* and not a 'third level' within the federal system. Nevertheless, they have their own responsibilities and a certain degree of independence. The Constitution stipulates in Article 28(2) that they must be given the opportunity 'to regulate all local affairs on their own responsibility, within the limits prescribed by the laws'.

Cooperative federalism

To ensure sufficient cooperation between the Federal and *Länder* levels on legislation as well as in other areas, various agreements have been concluded between the responsible divisions of the Federal and *Länder* ministries. Moreover, there are regular meetings of the *Länder* heads of government and with the Federal Chancellor, the Standing Conference of the Interior Ministers of the *Länder*, the Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder*, the Economic Policy Council, the Financial Planning Council, the Science Council, the Joint Science Conference and many other joint working groups. These regular meetings facilitate cooperation among the *Länder* and with the Federal government. For certain topics working groups are set up, for instance, at the Interior Ministers Conference, where experts of the various ministries work together. Decisions are made unanimously.

On the basis of these structures, a form of cooperative federalism has developed in Germany, as the administrative and political science literature largely agrees (Münch and Meerwald 2002; Scharpf 1976; Schubert and Klein 2006). The term ‘cooperative federalism’ refers to the fact that decisions are made by various decision-making levels working together. Cooperative federalism is associated with the aim of keeping differences between the member states as minimal as possible and striving for equivalent living conditions. In Germany, this aim is supported by constitutional law. Article 106 of the Constitution deals with the issue of maintaining equivalent living conditions throughout the Republic. This aim has led to a complicated system of revenue-sharing among the *Länder* and the Federation which is intended to reduce regional differences in living standards.

Various forms of cooperation between different state actors have given rise to overlapping jurisdictions and patterns of coordination, as well as formal and informal rights to be involved in decision-making. In this way, the many different levels of organisation and decision-making in the federal system are intertwined, both horizontally and vertically. As a result, it is often difficult to enact reforms in Germany, and in many cases it is hard to tell who is responsible for what decisions.

In recent years, some have proposed competitive federalism as an alternative to cooperative federalism. This model calls for the regions to have more autonomous decision-making authority in order to increase competition between regions. It is hardly surprising that this model is especially popular with those actors who do not profit from revenue-sharing among the *Länder*.

In fact, the *Länder* have lost influence in the field of law-making relative to the federal and European levels. In the large majority of policy fields, the federal or the

European level has the power to pass legislation. In consultations between the Federation and the *Länder*, the resulting reduction in *Länder* legislative power often leads to an unusual situation separate from the attempt to balance different interests. The *Länder* stress their sovereignty and constitutional status. Federal proposals are sometimes rejected as interference in the domestic affairs of the *Länder*. This is why the federal government often finds it necessary to seek allies among the *Länder* at an early stage in order to pursue common initiatives and interests.

5.2 Reforming federalism and anchoring comparative studies in the Constitution

The constitutional basis for benchmarking studies between Federal and *Land* administrations — Article 91(d) — is one element of the constitutional reforms achieved by the coalition of the two major parties, the CDU–CSU and the SPD, between 2005 and 2009 to modernise the federal system (phases 1 and 2 of the federalism reform).

These reforms were prepared, right down to complete drafts of proposed legislation, by high-ranking commissions of the *Bundestag* and *Bundesrat* in 2003–04 (First German Commission on Federal Reform) and 2007–09 (Second German Commission on Federal Reform).

The first phase was primarily concerned with untangling some of the highly intertwined decisions at federal and state level and with increasing the powers of the *Länder*. The second phase involved negotiating new constitutional and sub-constitutional financial legislation, in particular to limit government debt and budgetary emergencies. In addition to the finance reforms, the second phase also dealt with reforms in the area of administration. Here, two new articles were added to the Constitution: First, Article 91(c) of the Constitution governs IT cooperation between the federal and *Länder* levels. Secondly, the constitutional reform effective 1 August 2009 gave comparative studies constitutional status. The *Bundestag* and *Bundesrat* approved the new Article 91(d) of the Constitution worded as follows:

To determine and promote the productivity of their public administration, the Federation and *Länder* may conduct comparative studies and publish the results.

In analysing the debates over ‘benchmarking’, it is noticeable that in the commission’s meetings this issue found broad acceptance at the political level. The central argument was that comparative studies between administrative levels are especially suited to a federal system. Comparative studies are an appropriate instrument for federalism. A major advantage of federal systems over centralised state systems is competition for the best solutions. This advantage of federalism

only comes into play, however, when competition is organised, as it can be through comparative studies (see Fenna, this volume).

This position was partly based on the experience of often-significant differences in the way the different *Länder* carry out federal law — for example, with regard to agency organisation; administrative procedures; and the use of information technology. The result has been major differences in quality and costs — for example in the tax administration. This points to reserves of efficiency and effectiveness. Through administrative cooperation on comparative studies, individual reserves can be identified and necessary changes made. Comparative studies can also provide interesting insights for internal administrative services.

However, it was unclear what would provide the legal basis for comparative studies and what organisational form they would take. There was discussion as to whether legal provisions were even needed, or whether an intergovernmental agreement or a constitutional amendment would be the proper avenue. In terms of organisation, establishing a joint benchmarking agency was discussed.

Ultimately, the Commission on the Modernisation of Federation–*Länder* Financial Relations recommended to the *Bundestag* and *Bundesrat* that comparative studies be included in the Constitution. In its decision, the commission stated:

The commission agrees with the chairs that comparative studies of the public administration have proved to be a useful instrument. Comparative studies increase the transparency of state action and make it possible to recognise the best solutions and thereby optimize the public administration.

The *Bundestag* and *Bundesrat* also agreed with this position and adopted Article 91(d) of the Constitution. This article sends an extraordinary signal for administrative policy and is likely to have a binding effect on the Federal and *Länder* governments. It calls on the Federal and *Länder* administrations to conduct comparative studies.

5.3 A digression: experience with comparative studies

Even before the constitutional amendment, public administration in Germany had had experience with comparative studies, albeit, largely at the local level. There is much less experience at *Länder* level.

The local authorities network IKO of the local government association KGSt, where comparative studies groups at local level are coordinated, has the most experience.³

³ For more information, see: www.kgst.de.

The KGSt has overseen these projects, more than 240 of them since 1996, in areas such as facilities management, personnel management, youth welfare, public service offices and many more. More than 2600 municipalities and institutions have taken part.

At federal level, from 2000 to 2006 comparative studies were conducted with regard to grant procedures; IT and human resources processes; and in the area of sickness allowances for civil servants (Federal Ministry of the Interior 2005, p. 12). Also worth mentioning is the management of training conducted by the Federal Academy of Public Administration (BAköV) a central federal training facility (Federal Academy of Public Administration 2008), and comparison of services centres introduced as a pilot (Federal Ministry of the Interior 2009, p. 14). Additional comparative studies were carried out as part of the introduction of cost-results accounting (Federal Ministry of Finance 2007, p. 12).

Some comparative studies have also been carried out between *Länder*. In particular, the following should be mentioned:

- Benchmarking by the city-states of Berlin, Hamburg and Bremen on budgetary data based on finance policy statistics (Berlin Senate Department of Finance, Hamburg Department of Finance and Bremen Senator for Finance 2008).
- The international comparative studies Program for International Student Assessment (PISA), *Internationale Grundschul-Lese-Untersuchung* (IGLU) and ‘Trends in Mathematics and Science Study’ (TIMSS) have led to further *Länder* comparative studies at national level. It should be noted that the PISA-E studies represent a national supplement to the international PISA studies and are aimed at analysing the possible influence of external factors (for example, the school systems in the various *Länder*, class structure, level of parents’ education and provision of material goods at home) on pupils’ levels of achievement. In 2009, the PISA-E studies were replaced by studies comparing educational standards in the different *Länder* in order to determine whether pupils are attaining the standards set by the Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder*.
 - In addition, further comparative studies on specific issues are conducted, such as language acquisition (Köller, Knigge and Tesch 2010).
 - The PISA Consortium of Germany was founded for the PISA studies, which are conducted every three years (PISA-Konsortium Deutschland 2007). In order to ensure cooperation between the Federation and the *Länder* on comparative studies in the field of education, the following text was inserted into Article 91(b)(2):

-
- ... The Federation and the *Länder* may mutually agree to cooperate for the assessment of the performance of educational systems in international comparison and in drafting relevant reports and recommendations.
- Benchmarking for public procurement was launched under the title ‘REPROC-Excellence’. In the initial project phase, public contracting authorities such as the Federal Ministry of Economics and Technology are working with the German Association of Materials Management, Purchasing and Logistics and the Research Centre for Law and Management of Public Procurement at the University of Munich to develop performance-specific criteria for measuring public procurement.⁴
 - In 2009 and 2010, various *Länder* administrations initiated the projects ‘Housing Benefit Made Easy’ (Federal Chancellery / National Regulatory Control Council 2009a); ‘Parental Benefit Made Easy’ (Federal Chancellery / National Regulatory Control Council 2009b); and ‘Federal Higher Education Grants Made Easy’ (Federal Chancellery / National Regulatory Control Council 2010) and conducted them together with the Federal Government and the National Regulatory Control Council (NRCC). The NRCC is an advisory body instituted in 2006 that helps the Federal government reduce the administrative costs associated with legislation.
 - In 1998, Saxony and Bavaria initiated a comparative study of tax offices, in which seven of the 16 *Länder* are now participating.⁵

The examples of comparative studies in the finance administration and the project ‘Federal Higher Education Grants Made Easy’ will be used to explain briefly and in greatly simplified form how comparative studies can be carried out.

The studies initiated by Saxony and Bavaria in 1998 began by defining and comparing indicators for the following areas: task completion; client and staff satisfaction; and cost-effectiveness. Based on these indicators, reports on the performance of the tax offices were produced. In addition, more in-depth analyses, such as business process analyses, were carried out to explain differences between the tax offices. These reports and analyses provided the basis for adopting management targets and measures to optimise the respective administrations. The *Länder* administrations managed the comparative studies and modernisation process on a collegial and equal footing. In particular, contracts between the responsible actors and ongoing reporting requirements were used. Overall, the sharing of information served the ongoing improvement of participating tax offices. According

⁴ See: <http://www.bme.de/REPROC-Excellence.reproc-excellence.0.html>.

⁵ See: <http://www.leistungsvergleich.de/index.html>.

to those involved, these comparative studies have helped bring about significant improvements in quality as well as client and staff satisfaction in the tax offices.

In addition to the Federal government, the administrations of eight *Länder* are taking part in the project to simplify grants for students in higher education. This project was initiated by the *Länder*, and a process was chosen which is coordinated with all stakeholders. The aim was to identify administrative burdens and differences in implementation with regard to the making and processing of applications for educational grants and to simplify measures and/or develop services to reduce the burden on students and grant offices. First, the burden for applicants and the responsible administrations was identified, including the analysis of sub-processes. Based on these indicators, which vary by organisation, more in-depth analyses were conducted (for example, surveys, analyses of business processes). In this way, it was possible to identify differences specific to certain *Länder* and offices and to recommend improvements. Recommendations covered the design of application forms, methods for checking plausibility and errors, and the possibility of online application, to mention just a few.

In addition to these projects in which the various governments were directly involved, non-governmental organisations also carried out benchmarking projects. One example is the consumer protection index first commissioned by the Federation of German Consumer Organisations in 2004 and updated every two years since then. The index presents a comparative analysis of the consumer policy profile of the *Länder* (Bridges 2010). According to the Federation of German Consumer Organisations, this comparison has led to major efforts and improvements in the consumer protection policy of the *Länder*.

5.4 Activities to implement Article 91(d) of the Constitution

The constitutional amendment sends an important administrative policy signal. In this way, the *Bundestag* and *Bundesrat* have given modernisation of public administration constitutional status. This prominent position should be understood in the context of government deficits, changing expectations of business and individuals and international competition of entire regions. Performance of public administration and the quality of public services have long been important factors for investors and entrepreneurs when choosing where to do business.

The parties of the governing coalition (CDU/CSU and FDP) have recognised the need for better performance by public administration and included in its coalition agreement for the 17th legislative term a call to conduct comparative studies. Comparative studies

must become an instrument of administrative development. Areas in which comparative studies are to be conducted are to be defined in an annual work program.

The government program ‘Transparent and Network-Based Administration’ adopted by the Federal Government on 18 August 2010 includes the statements that ‘subjects for comparative studies should be designated in an annual work program’ and ‘that every ministry should take part in at least one comparison group by 2013, if possible’ (The Federal Government 2010, p. 52). Pilots on health management and balancing family and career are currently in planning. It is not yet clear how these activities can be expanded to the desired extent.

Situation at the federal level

Working Group VI of the Standing Conference of Interior Ministers, whose responsibilities include administrative organisation, drafted a plan for conducting comparative studies in public administration. This plan discusses the aims of comparative studies, how to go about them, how to handle their results, and the role of the Standing Conference. However, it does not say anything about the organisational structures in which comparative studies should be carried out.

The plan was approved by the Standing Conference of Interior Ministers on 18-19 November 2010. The Standing Conference also agreed that, as an initial step, each of its working groups should designate a specific subject and participants for a comparative study ahead of their spring 2011 session. With these decisions, the Standing Conference has decided to begin implementing Article 91(d) of the Constitution. It remains to be seen which specific subjects for comparative study will be designated, who will participate in the projects and how the comparative studies will be carried out.

As early as 2004, the heads of the *Länder* governments spoke out in favour of more comparative studies. In the context of Article 91(d) of the Constitution, for their session on 15 December 2010 they asked for reports on the ongoing efforts of all the conferences of specialised ministers regarding comparative studies. This comprehensive overview will provide further information about the implementation of the constitutional amendment.

In its 2010 annual report, the National Regulatory Control Council recommends applying the lessons learnt from the projects ‘Housing Benefit Made Easy’, ‘Parental Benefit Made Easy’, and ‘Federal Higher Education Grants Made Easy’ for comparative studies in accordance with Article 91(d) and offers its assistance (National Regulatory Control Council 2010, p. 40).

With these activities in mind, it can be expected that further comparative studies in various policy fields of the federal system will be launched soon and that in the medium term comparative studies will be organisationally anchored for the purpose of policy and administrative coordination.

References

Berlin Senate Department of Finance, Hamburg Department of Finance and Bremen Senator for Finance 2008, *Benchmarking der Stadtstaaten*, Berlin, Hamburg and Bremen.

BRIDGES – Politik und Organisationsberatung 2010, *Das verbraucherpolitische Profil der Länder im Vergleich: Abschlussbericht zum vierten Verbraucherschutzindex der Bundesländer*.

Federal Academy of Public Administration (ed.) 2008, *Bildungscontrolling in der Bundesverwaltung* (Controlling of training in the federal administration).

Federal Chancellery / National Regulatory Control Council 2009a, *Einfacher zum Wohngeld* (Housing Benefit Made Easy).

— 2009b, *Einfacher zum Elterngeld* (Parental Benefit Made Easy).

— 2010, *Einfacher zum Studierenden-BAföG*, (Federal Higher Education Grants Made Easy).

Federal Ministry of Finance 2007, *Einführungsstand der Kosten- und Leistungsrechnung in der Bundesverwaltung: Fortschrittsbericht an den Rechnungsprüfungsausschuss des Deutschen Bundestages* (Status of the introduction of cost-results accounting in the federal administration: Progress Report for the Parliamentary Public Accounts Committee of the German Bundestag).

Federal Ministry of the Interior 2005, *Progress Report of the Government Program 'Modern State – Modern Administration' in the Field of Modern Administrative Management*.

— 2009, *Umsetzungsplan 2009: Fortschrittsbericht zum Regierungsprogramm „Zukunftsorientierte Verwaltung durch Innovationen“* (Implementation Plan, progress report on the government program 'Focused on the Future: Innovations for Administration').

Köller, O., Knigge, M. and Tesch, B. 2010, *Sprachliche Kompetenzen im Ländervergleich*.

Münc, U. and Meerwaldt, K. 2002, 'Politikverflechtung im kooperativen Föderalismus', *Informationen zur politischen Bildung*, vol. 275.

National Regulatory Control Council 2010, *Qualität durch Transparenz: Mit Bürokratieabbau zu moderner Gesetzgebung* (Quality through transparency: Reducing bureaucracy to modernize legislation).

PISA-Konsortium Deutschland (ed.) 2007, *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*, Münster, 2006.

Scharpf, F.W., Reissert, B. and Schnabel, F. (eds) 1976, *Politikverflechtung, vol. 1: Theorie und Empirie des kooperativen Föderalismus in der Bundesrepublik*, Kronberg/Ts.

Schubert, K. and Klein, M. 2006, *Das Politiklexikon*, 4th edn, Bonn.

The Federal Government 2010, *Regierungsprogramm, Vernetzte und transparente Verwaltung* (The government program 'Transparent and network-based administration').

6 Benchmarking sustainable development in the Swiss confederation*

Daniel Wachter¹

Swiss Federal Office for Spatial Development

Over the past decade, Switzerland has developed a collaborative system of intergovernmental benchmarking to promote sustainable development across the country. It is a voluntary arrangement, wherein participating Cantons (states) and municipalities report on an agreed range of performance indicators and the full results are made public. In this system an agency of the federal government — the Federal Office for Spatial Development — plays a facilitative and coordinating but not directing role. Over time, the system has proven successful in attracting participation from more and more Cantons and municipalities and in having its findings incorporated into policy making processes. A good part of its success can be attributed to the highly collaborative and consensual way in which it has developed, an outcome that reflects the realities of Swiss federalism and concurrent nature of responsibility in this area.

6.1 Swiss federalism

The *Confœderatio Helvetica*, or Swiss Confederation, is the oldest and most decentralised federation in world. With fewer than 8 million people, it is made up of 26 Cantons and has three national languages.² Under the Constitution (Article 3), ‘the Cantons are sovereign except to the extent that their sovereignty is limited by the Federal Constitution. They shall exercise all rights that are not vested in the

* Daniel Wachter did not present at the Melbourne roundtable. This chapter reflects his contributions to other Forum benchmarking events, with themes closely related to those of the roundtable.

¹ Daniel Wachter is Head, Sustainable Development Section, Federal Office for Spatial Development (Bundesamt für Raumentwicklung, or ‘ARE’), CH-3003 Bern, Switzerland.

² 64 per cent of the population speak German; 21 per cent speak French; 6 per cent speak Italian; 9 per cent speak another language.

Confederation.’ With a strong sense of identity and a strong tax base, the Cantons continue to be major players in the federation, insisting on their independence and rejecting direction from the federal government (Linder 2010). Municipal government also has a well-established place in the Swiss political system, and, like the Cantons, is largely self-financing.

The imposition of programs by the federal government on the Cantons or municipalities is thus not a characteristic part of Swiss federalism as, for instance, it is in Australia or the United States. In the case of sustainability policy, coordinated action reflects constitutional requirements. Article 2 (‘Aims’) of the Constitution states that sustainable development is a national objective; and Article 73 (‘Sustainability’) makes environmental protection a mandatory criterion of policy. ‘The Confederation and the Cantons shall endeavour to achieve a balanced and sustainable relationship between nature and its capacity to renew itself and the demands placed on it by the population’. Since many of the substantive matters relevant to sustainability fall within the jurisdiction of the Cantons, this constitutional task falls as much to the Cantons as to the Confederation to execute and is thus a concurrent responsibility.

6.2 A collaborative framework

This program of sustainability benchmarking is open to all the Cantons and the cities that are ready to commit the necessary resources. The group of participants has grown steadily over the last few years, reaching 19 Cantons and 16 cities at the end of 2011 (figure 1). It is a classic case of what Fenna (this volume) describes as ‘external’ or ‘collegial’ benchmarking. That is to say, it is a voluntary exercise where the central government plays a strictly facilitating and moderating role and no sanctions of any form are involved. The federal agency’s facilitative role is particularly in respect of providing the technical foundation for benchmarking.

The participating Cantons, cities and federal offices together form the sponsors of the *Cercle indicateurs*, or ‘indicators group’ — ‘a forum dedicated to the development and use of sustainability indicators for Swiss cities and Cantons’. The Federal Office for Spatial Development is responsible for managing the project. Initially, a private consulting firm oversaw the project office and all the technical issues concerning data collection and management. Since 2008, the Federal Statistical Office has been a partner in the project and is responsible for managing the project office and the data. Through this change in leadership of the project office from a private consulting firm to the Federal Statistical Office, it was possible to reinforce the legitimacy of the data through greater quality control. Today, the

indicators and the operation of the indicator system correspond to a large extent to the requirements of official statistics.

Participation in the *Cercle indicateurs* is voluntary and essentially open to all Cantons and cities; there is no legal requirement or pressure to participate. Over the last few years, because of the growing number of participants and the stricter requirements regarding the quality of the indicators, managing the project has become increasingly labour-intensive — so much so that the financing of a project office by the participants became necessary. To do this, the Cantons and the cities had to commit for the first time to a long-term contract for the period of 2010–13. Participation continues to be voluntary, but it is linked to sharing the project costs and a multi-year commitment. Whoever signs the contract with the Federal Office for Spatial Development declares their support for the quality charter for official statistics (Eurostat 2005; Federal Statistical Office and Konferenz der Regionalen Statistischen Ämter der Schweiz Korstat 2002) and also states that they agree to the collected data being published.

There are no direct political consequences for participants in the *Cercle indicateurs*. It is purely a monitoring system and not part of the control or management systems of a higher-ranking office and thus has no reward, penalty or sanction mechanisms at its disposal. A poor performance at the *Cercle indicateurs* level does not result in reduced subsidies or any other punitive measure. This promotes largely unrestricted participation in the discussions and exchanges relating to experience.

A statute setting clear rules of the game

A statute agreed to by all sets out clear rules regarding collaboration within the *Cercle indicateurs*. As well as the aims, administration and processes of the collaboration, the statute governs, in particular, the following points:

Meetings and resolutions

The sponsors are to meet at least twice a year. Working groups may be formed as needed, to which the sponsors can also delegate decision-making powers (for example, in connection with a review of the indicators). As a matter of principle, a consensus should be strived for when making decisions. If this proves impossible, the decision by the majority shall apply. Each participating federal agency, canton and city has one vote.

Significance of official statistics

In principle, the quality criteria of official statistics must be observed when selecting and defining indicators. All the statistical activities of the *Cercle indicateurs* must comply with the production and distribution standards of official statistics. This relates in particular to the independence, impartiality and quality of the data, data security, and the publication of the statistical results.

Responsibilities

In addition to organising surveys, including exact instructions aimed at the Cantons and cities on how to uniformly record decentralized indicators and publish all the results on the internet, there is also—with regard to obtaining the best quality and comparison possible—the periodic check and review, if and when necessary, of the indicators. Discussing experience gained concerning the application of the results counts as one of the responsibilities.

Periodicity

The Cantons collect data every two years and the cities every four. The Federal Statistical Office also makes centralised data available to the cities every two years.

Products

The products of the *Cercle indicateurs* are defined as:

- (a) the original values of the indicators
- (b) a profile of the strengths and weaknesses of each participant (expressed as utility values)
- (c) a graphic representation for each participant of the deviation from the mean of the utility values
- (d) a comparison with the other Cantons or cities, respectively for each indicator (expressed as original values)
- (e) aggregated benchmarking, on the one hand, with a total value for each sustainability area and, on the other hand, a comprehensive total value.

Financing

The Cantons share the costs by each one paying one standard amount representing a total of approximately 50 per cent of the costs, while the cities pay approximately

20 per cent of the costs depending on population. The remaining costs are covered by the participating federal agencies.

Responsibility and entrenchment in the cantons and cities

Since the *Cercle indicateurs* is not a sectoral-political system, responsibility and entrenchment within the Cantons and cities will be managed differently. The contract partners with the Federal Office for Spatial Development, for example, are government departments, cantonal chancelleries, city councils or the agencies responsible for the environment, spatial development, the economy and statistics. The representatives among the sponsors are the officials of the various agencies, mostly from the areas of the environment, spatial development, the economy or statistics. Various Cantons and cities are each represented by two people, most of them by a specialist in one of the three key areas of sustainable development and by one statistics expert.

6.3 Sustainable development and the *Cercle indicateurs*

The Swiss approach to sustainable development seeks to address major environmental, economic and social challenges. Domestically, these relate most importantly to the unsustainable ecological footprint of modern industrial society — with the *per capita* level of the Swiss population 200 per cent higher than can be maintained globally over the long term. At the same time, sustainable development also involves a commitment to meeting the economic and social needs of the world's population. This requires long-term, fundamental, economic and social structural change.

Sustainable development is often illustrated using three circles or pillars to represent the key areas of the environment, the economy and society. On the one hand, this is to illustrate the link between economic, social and ecological processes and, on the other hand, that negotiations among public as well as private stakeholders should not occur in an isolated and one-dimensional manner, but rather that they take into account the interplay between on the three key areas and its impact. The measures developed by the *Cercle indicateurs* also follow this structural principle (table 6.1).

The indicator sets consist of approximately 35 indicators that cover the areas of environment, economy, and society. On the one hand, some of these are taken from the official statistics of Switzerland (the so-called centralised indicators), while, on

the other hand, some are those that must be collected by the participants themselves (decentralised indicators).

Figure 6.1 Overview of the group of participants in 2011



Source: Federal Office for Spatial Development

Since 2000, various Cantons and cities have been hard at work defining a cantonal and a municipal indicator system for sustainable development. In total, there were five bottom-up indicator initiatives with several participants for the most part (*Cercle indicateurs* 2005, p. 4). The *Cercle indicateurs* was created in 2003 out of these various bottom-up initiatives. In 2003, the Federal Office for Spatial Development switched to coordinating the various cantonal and municipal initiatives to produce uniform and, therefore, comparable indicator systems for Cantons and cities.

Goals of the *Cercle indicateurs*

The *Cercle indicateurs* program pursues the following goals (*Cercle indicateurs* 2005, p. 4):

-
- Determination of a consensus-building indicator system and periodical data collection — this presumes the availability of indicators capable of building a consensus. To do this, all corresponding data must be made available to the participants, and the majority of the participants must speak out in favour of these indicators. This also includes the periodic collection of data.
 - Monitoring and benchmarking of sustainable development at the strategic level of general policies — on the one hand, the data collected help the participants to observe their own development over time (monitoring) and, on the other hand, they help them to draw comparisons with other participants (benchmarking). The highly aggregated indicator set covering a number of sectoral issues provides information on sustainable development primarily at the strategic level of general policies and not individual sectoral policies.
 - Enhanced indicator content — the basic data may change, new data may become available, or data used to date can no longer be collected. Moreover, as the indicators were applied, deficiencies in some indicators were discovered which meant that the affected indicators had to be adapted. Therefore, a systematic, regular review of the basics must be performed and a discussion held by the sponsors about the possibility of making improvements.
 - Platform to exchange ideas on how to apply the indicators, for example, for reporting on sustainable development — the *Cercle indicateurs* does not specify how the Cantons and cities should use the indicators; however, it does serve as a platform for exchanging ideas on a variety of topics relating to the indicators such as data collection, outcome controls and the application of the indicators.

Misunderstandings always arise with respect to indicator systems when their goal in relation to the three-sided concept of monitoring–controlling–evaluation is unclear. The concept of monitoring refers to constant observation. In this way, problematic developments can be detected early. The concept of controlling is rooted in business administration where it is an instrument of a goal-oriented business management defined within the closed loop of planning, translation, control and (counter-) managing. A continuous and documented analysis of the deviation from goals within a reporting system forms the foundation for measures to counter-manage. Evaluation is defined to a large extent as a scientific and empirically supported judgment of the concept, the execution and/or the effectiveness of state activities. Evaluations assess state activities based on transparent criteria and present cause and effect relationships between the activities and the impact. This kind of information can help with decision-making, financial reporting and controls, or serve as the basis for qualified discussions.

The *Cercle indicateurs* program is clearly aimed at monitoring, that is, it is about the observation of sustainable development. The indicators were not designed to

control policy, that is, they are not about performance management; nor do they directly serve policy assessment goals. Instead they form, at the most, a basis for raising issues (the ‘can opener’ role). The participating Cantons and cities are free to decide on how the indicators are applied. That said, some of them are moving entirely in the direction of applying them to policy control.

More clarity is needed to transfer the concept of ‘sustainable development’ to the political realm. Indicators correspond to this kind of reality and help to provide an ongoing assessment of the situation. Within the framework of the *Cercle indicateurs*, a set of so-called core indicators was defined to allow the Cantons and cities to assess a situation in terms of sustainable development. The core indicators are those indicators that describe the central elements of sustainable development for the entire system and are chosen by participating Cantons and cities for each corresponding level. In concrete terms, this was about developing one indicator system for Cantons and another one for cities that makes sense at the cantonal and the municipal level, respectively. This objective requires that easily understood measurements be selected when choosing the core indicators and that they allow as much room to manoeuvre as possible for each canton or city.

A tool of a broader governance system for sustainable development

The *Cercle indicateurs* is one part of a comprehensive system of knowledge-based approaches used today in Swiss policies on sustainable development (Wachter 2007; 2010). It consists of instrumental approaches to observe sustainable development using indicators — the *Cercle indicateurs* at the cantonal and municipal levels complements MONET,³ the monitoring system for sustainable development at the national level (Federal Statistical Office *et al.* 2003a) — at the national as well as the subnational level, to differentiate mostly qualitative decisions regarding projects from the point of view of sustainable development (Federal Office for Spatial Development 2008) as well as the management and periodical political assessment of the translation of the Federal Council’s sustainable development strategy.

In addition, the *Cercle indicateurs* is part — or a project — of a broader sustainable development arrangement between the confederation, the Cantons and the municipalities: the so called ‘Forum for Sustainable Development’ is a vertical co-ordination and exchange platform focusing on policy issues related to sustainable development with regular plenary meetings and a number of ancillary activities, like among others the *Cercle indicateurs*. The Forum is the place where peer review approaches are practiced — when, for example, a canton reviews its

³ ‘Monitoring der nachhaltigen Entwicklung’ (Monitoring Sustainable Development).

sustainable development strategy and invites representatives of other Cantons to comment and give advice.

6.4 Operationalisation

The core indicators of sustainable development have the following characteristics (*Cercle indicateurs* 2005, p. 2):

- Objective — monitoring sustainable development (raise awareness, present strengths and weaknesses, assess a situation, indicate development trends)
- Geopolitical reference — the political limits of a canton (and not a region) or a city (and not a metropolitan region)
- Decision-making levels — general policies, not individual concepts or projects
- Content orientation — overall, not oriented towards individual areas of expertise
- Scope — a manageable number of indicators that are easily communicated.

The core indicator system is aimed at the three key areas concept of sustainable development introduced earlier. It makes the three key areas real by using 35 so-called ‘target areas’ that are each measured as a rule by one indicator for the Cantons and one for the cities (table 6.1). Depending on the complexity of the target area and the data available, no core indicator or two core indicators will be set for individual target areas.

Table 6.1 Overview of the target areas and core indicators

<i>Target area</i>	<i>Cantonal core indicator</i>	<i>Municipal core indicator</i>
ENV1: Biodiversity	Cantonal breeding bird Index (place holder)	Municipal breeding bird index
ENV2: Nature and Landscape	Surface area of valuable natural spaces	Surface area of valuable natural spaces
ENV3: Energy Quality	Renewable energy, incl. waste heat (place holder)	Renewable energy, incl. waste heat (place holder)
ENV4: Energy Consumption	Total energy consumption (place holder)	Electricity consumption
ENV5: Climate	CO ₂ emissions (place holder)	CO ₂ emissions (place holder)
ENV6: Raw Material Use	Amount of waste per inhabitant	Amount of waste per inhabitant
ENV6: Raw Material Use	Sorted collection rate	Sorted collection rate
ENV7: Water Balance	Water discharge via waste water purification facility	Water discharge via waste water purification facility
ENV8: Water Quality	Nitrates in the ground water	Transport of effluent from the waste water purification facility
ENV9: Land Use	Built-up areas	Built-up areas
ENV10: Soil Quality	Heavy metal contamination of land (place holder)	No indicator
ENV11: Air Quality	Long-term pollution Index	PM10 Emissions (place holder)

(continued next page)

Table 6.1 (continued)

<i>Target area</i>	<i>Cantonal core indicator</i>	<i>Municipal core indicator</i>
Key Area: Economy		
ECON1: Income	Cantonal aggregate income	Taxable income of individuals
ECON2: Cost of Living	Rental price level	Rental price level
ECON3: Labour Market	Rate of unemployment	Rate of unemployment
ECON4: Investments	Renovation and maintenance costs	Renovation and maintenance costs
ECON5: True Costs	No indicator	Application of the polluter pays principle
ECON6: Resource Efficiency	No indicator	No indicator
ECON7: Innovation	Employees in innovative fields	Employees in innovative fields
ECON8: Economic Structure	Employees in high value-added industries	Employees in high value-added industries
ECON9: Know-how	Qualification level	Qualification level
ECON10: Budget	Health of cantonal finances	Health of municipal finances
ECON11: Taxes	Tax burden index	Tax burden of individuals
ECON12: Production	No indicator	No indicator
Key Area: Society		
SOC1: Noise/Quality of Housing	Impact of traffic noise	Traffic calming zones
SOC2: Mobility	Access to public transit	Access to public transit
SOC3: Health	Potential lost years of life	Potential lost years of life
SOC4: Security	Road traffic accidents with personal injury	Road traffic accidents with personal injury
SOC4: Security	Violent offences	Criminal charges
SOC5: Income/Wealth Distribution	Low-income taxpayers	Gini Coefficient for income distribution
SOC6: Participation	Voting and polling	Voting and polling
SOC7: Culture and Recreation	Cultural and recreational expenses	Cultural and recreational expenses
SOC8: Education	Youth education	Broken educational thread
SOC9: Social assistance	Access to social assistance services	Access to social assistance services
SOC10: Integration	Naturalisation of immigrants	Naturalisation of immigrants
SOC11: Equal Opportunity	Women in management positions	Number of daycare spaces
SOC12: Supraregional Solidarity	Relief operations	Relief operations

Source: Federal Office for Spatial Development.

The proposed core indicators were selected from a range of possible indicators based on the following criteria. They had to:

- be as representative and as meaningful as possible for the target area
- be quantifiable
- be easy to understand and to communicate

-
- represent the greatest possible consensus among those participating in the process
 - be capable of being influenced as a general rule by municipal and cantonal authorities.

It should be added regarding the last point that the *Cercle indicateurs* — in contrast to MONET, a much more extensive and detailed monitoring system of sustainable development on the national level (Federal Statistical Office *et al.* 2003b) — does not have access to a sophisticated typology of indicators concerning circumstances, resources, production rates, the source of problems or political measures. Instead, there is a simple underlying policy outcome model that differentiates between policy inputs (money or other resources), outputs (specific services or products), impact (effect on target groups) and outcomes (effects at the end of the causal chain). Indicators are always being sought for outcomes, even when this is not always possible due to limited data availability.

Data collection

The centralised indicators rely primarily on the official statistics of the federal government and are provided by the Federal Statistical Office. A few centralised indicators rely on data that are purchased or collected by data suppliers outside the official statistical system. The decentralised data must be collected by the Cantons and the cities themselves — with the help of precise instructions from the Federal Statistical Office — and reported to the Federal Statistical Office within the deadline. The accuracy of the data must be checked by the participating Cantons and cities. The Cantons must collect the data every two years and the cities must collect the data every four years. The Federal Statistical Office also makes the centralised data available to the cities every two years, as needed.

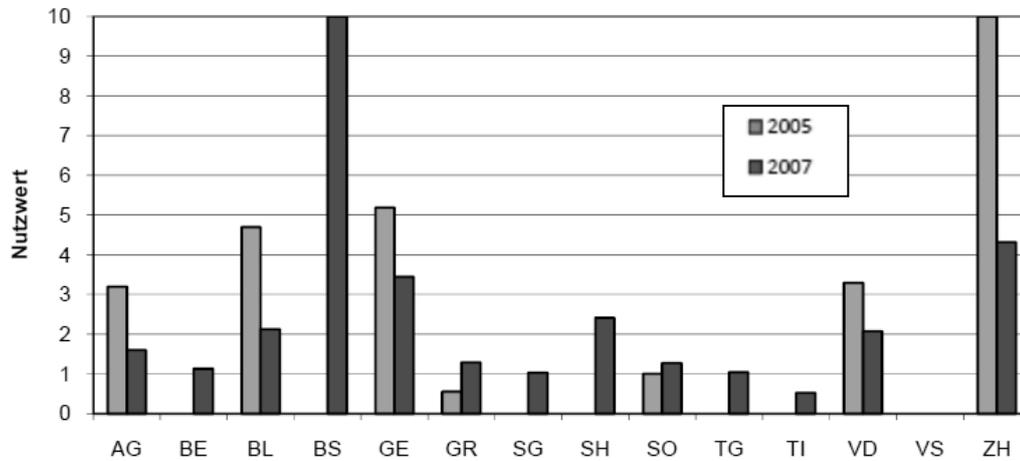
For cost reasons, as a general rule *Cercle indicateurs* relies on existing data sources. Inevitably this sometimes entails compromises in the selection of indicators. However, where the data situation has been truly unsatisfactory, *Cercle indicateurs* has selected place-holders, signalling to the relevant bodies, such as statistical offices, a need to generate new data. In certain cases, such as ENV2 (surface area of valuable natural spaces), *Cercle indicateurs* itself not only defined the indicator, but also organised the generation of corresponding data.

The benchmarking principle

At the start of the *Cercle indicateurs* program, the actual indicators were converted to typical utility values for the purpose of comparison and also to calculate the mathematical aggregation. The values of all participants were applied proportionately to a scale of zero to ten, with the ‘best’ participant receiving a ten and the ‘worst’ participant receiving a zero. There were two deciding factors. First, the change in the indicator (increase or decrease), indicating the direction of sustainable development, can be determined for the 35 target areas of the *Cercle indicateurs*; however, in most cases there are no concrete normative target values. Only the ‘best’ and the ‘worst’ can be indicated among the recorded participant values. Secondly, the conversion of actual values into utility values makes it possible to compare indicators using different measurements, such as amount of money, or units of weight or square measurements. Moreover, the utility values also allow for a mathematical aggregation of indicator values in the three key areas or as a total value, which makes it possible to prepare a ranking ideal for communication purposes.

This original benchmarking principle certainly had an annoying disadvantage in connection with the —fortunately— growing group of participants that would change from survey to survey: the ranking of a participant with respect to an indicator or an aggregate could vary greatly because of the addition of a new participant, even if the actual values had hardly changed. This impact is shown in figure 2. Between the surveys of 2005 and 2007, the Cantons of Basel-City (BS), St. Gallen (SG), Schaffhausen (SH), Thurgau (TG), Ticino (TI) and Valais (VS) joined as new participants. In the 2005 survey, the canton of Zurich, for example, had the highest value for cantonal aggregate income per capita, for which it received the utility value of ten. In 2007, another canton was added when Basel-City (BS) joined, which, because of the many corporate headquarters there, also has an exceptionally high aggregate income compared with Zurich (ZH). Zurich’s (ZH) utility value dropped to approximately four in 2007 only because of the new participant, without its own aggregate income changing appreciably in any way. These kinds of jumps are highly misleading and difficult to explain to the public.

Figure 6.2 Typical benchmarking using the example of the indicator ‘cantonal aggregate income per capita’

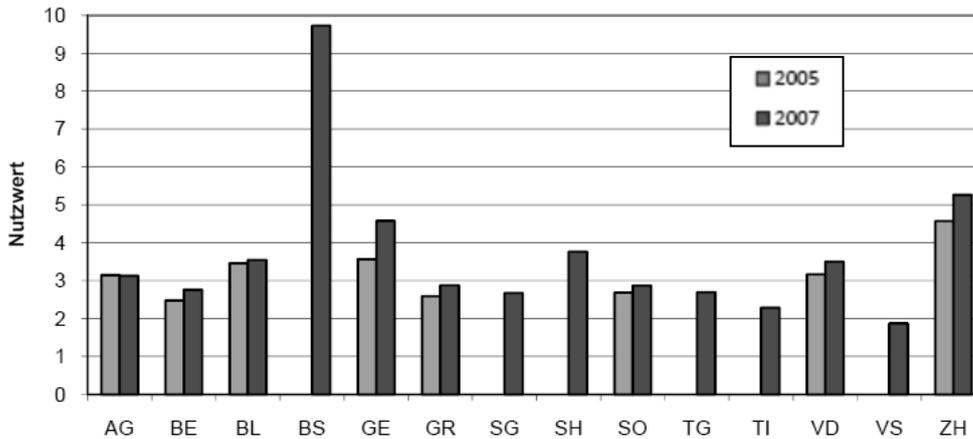


Nutzwert = Utility value.

To address this problem, a new utility value principle was introduced in the 2009 survey (using 2007 data). The end points of zero and ten no longer represent the ‘worst’ and the ‘best’ value. Rather, an absolute scale was established. For every indicator, the sponsors decided on a range for which the end points were not to represent normative target values, but rather help to calculate the utility values only. The lower and upper limits were established based on the values of earlier surveys, so that the anticipated values for the next 20 years or so could be mapped out. In the example of the cantonal aggregate income, the lower limit of CHF 20 000 (with a utility value of zero) and the upper limit of CHF 110 000 (with a utility value of ten) were used to set the range. Given that the utility values will now be based on an absolute scale, it will be possible to compare the utility values over time even when the group of participants changes. Also, it will still be possible to compare different indicators and calculate the mathematical aggregation.

Figure 3 shows the dramatic consequences of the new benchmarking principle. The utility value of the canton of Zurich (ZH) is no longer falling in 2007 because of the new participant Basel-City (BS), rather its value is rising because the aggregate income actually rose.

Figure 6.3 **New benchmarking using the example of ‘cantonal aggregate income per capita’**



Nutzwert = Utility value.

Products of the *Cercle indicateurs*

The products of the survey consist of the original values (values in the specific unit of measurement); a profile of strengths and weaknesses (in utility values); a graphic representation of the deviation from the mean of the utility values; as well as a comparison with other Cantons and cities, respectively, for each indicator (in original values). These products, along with the meta-data (indicator definitions and other background information), are published on the website of the Federal Statistical Office.

Another product of the *Cercle indicateurs* is aggregated benchmarking, that is, the mathematical sum of a total value per key area (the environment, the economy, society) and of a total value overall. For the aggregation for each key area, all indicators for each key area are weighted equally. For the total aggregation the three key areas are weighted equally, which do not have the same number of indicators because of individual indicator gaps and a few target areas that have more than one indicator.

While the members of the *Cercle indicateurs* must participate in the collection and publication of the data, participation in aggregated benchmarking is voluntary. The result is not published on the website of the Federal Statistical Office, but rather on that of the Federal Office for Spatial Development. This situation is based on a certain ambivalence in relation to the aggregation of the indicator values and the preparation of a list ranking the Cantons and the cities. On the one hand, the aggregation has indisputable communications advantages in that the results of the *Cercle indicateurs* appeal to the media interested in simplification and pithy

slogans. And, in fact, in each case the publication of the aggregated benchmarking generates a lot of media attention. On the other hand, balancing ‘apples’ and ‘oranges’ invites a justifiable scepticism about the methodological integrity of the exercise. This is also reinforced in that the *Cercle indicateurs*, despite the applicable aim, cannot always determine the outcome indicators that would illustrate the results of political negotiations, so that the *Cercle indicateurs* does not exclusively show political achievements, but rather, at least, partly structural characteristics, such as an urban or rural situation.

In discussing and weighing the pros and cons of aggregated benchmarking, the sponsors decided on the following course of action:

- With every survey, the participants can voluntarily decide — only prior to the publication of the results, though — whether or not they want to participate in the aggregate benchmarking. Usually, two-thirds of the participating Cantons and cities do.
- Because of the questionable methodology set against the backdrop of the quality requirements of official statistics, the results of the aggregate benchmarking are not published on the website of the Federal Statistical Office, but rather on that of the Federal Office for Spatial Development, which is better able to weigh the different criteria of political and communications-related considerations versus statistical considerations.

6.5 Results and effects

As noted above, the *Cercle indicateurs* program concentrates on the definition and periodical survey of indicators as a monitoring/benchmarking exercise only, and does not actively interfere, as a Swiss national institution, in the way in which the indicators are used by the Cantons and cities. This restraint with respect to the objective corresponds to a conscious agreement reached by the participants. It also relates back to the origin of the *Cercle indicateurs* as a bottom-up initiative in which the federal government gradually assumed a largely moderating role. The program is, in other words, not part of a political control or management system. However, the sole objective of monitoring can also be attributed to the limited resources of those taking part, which negated the need to formulate a broader goal. As was already explained, the federal government provides the data, including the decentralised data provided by the Cantons and cities, to all participants. Use of the data is the exclusive responsibility of the participants. Nonetheless, the reality is that that the *Cercle indicateurs* wants to influence how the results are politically applied and, in the end, have an impact ‘on the ground’.

To date, the *Cercle indicateurs* had the following impact on the target groups:

- Increased number of participants — the fact that, in this voluntary benchmarking system, the number of participants rose from eight Cantons and 14 cities in 2005 to 19 Cantons and 16 cities in 2011 can be interpreted in such a way that the Cantons and the cities see this as a valuable system they can reasonably use for one of the following applications.
- Use of data as a basis for analysis — the indicators have many applications as the starting point for deeper analyses of individual problem areas and as the basis for formulating proposals for political negotiations.
- Reporting on sustainability — by 2011, eight Cantons (Aargau, Basel-City, Bern, Geneva, Schaffhausen, St. Gallen, Vaud and Zurich) and two cities (Baden and Zurich) prepared reports on the development of their jurisdiction and installed regular sustainability reporting on the basis of the indicators of the *Cercle indicateurs*.
- Use of the data for government/legislative programs — several Cantons and cities use the reports on sustainability as the basis for the medium- and long-term planning of responsibilities within the framework of government or legislative planning. They implement the indicators as the guiding principle at a political and a strategic level together with the New Public Management.
- Basis for, and adoption of, a sustainability strategy — many Cantons and cities use the *Cercle indicateurs* or, more precisely, the analytical fundamentals that arise, to adopt a broader sustainability political action program (Local Agenda 21 or similar). Provided they are already committed in this regard, they use the *Cercle indicateurs* to monitor progress.

What effect the *Cercle indicateurs* ultimately has at the outcome level is hard to determine in a system that is limited to monitoring objectives and is not part of a policy management mechanism; and is established at a overall political–sectoral meta-level. In addition, effects only occur over longer causal chains, in which the *Cercle indicateurs* assists by initiating or supporting cantonal or municipal sustainability programs, or by contributing to a more coherent and stronger goal-oriented policy by influencing the New Public Management. This also goes along with a long time delay until the effect of the outcomes becomes evident. All the same, we can see from the fact that the — voluntary — group of participants steadily grew, that the Cantons and cities involved in the *Cercle indicateurs* see the value of those longer-term contributory effects.

6.6 Lessons learnt

The following experience, findings and recommendations can be deduced from the roughly ten years the *Cercle indicateurs* program has been building, gradually continuing to develop and to operate.

Clearly defined goals, frames of reference and rules are essential

As research on the indicators shows (see, for example, Interface 2010), unclear goals for indicator projects always lead to conflict. The *Cercle indicateurs* benefited from the fact that its goals were clearly defined from the start. Also significant was the explicit definition of a conceptual frame of reference, which, in the case of the *Cercle indicateurs*, was monitoring. Successful cooperation among the stakeholders of all three government levels, for whom there was no pressure and no legal requirement to participate, is in no way guaranteed. While the first preliminary projects of the *Cercle indicateurs* involved small groups or even individual Cantons and cities, the likelihood of a conflict relating to the objective caused by different visions increased because of the increasing number of participants and their growing closer in a common enterprise. Uniting to set the rules and writing them in a statute proved to be extremely helpful and stabilising for the group. On the basis of this experience, the principle of establishing goals, frames of reference and rules is recommended in the case of indicator projects.

A participative approach ensures permanent support

In the case of the voluntary *Cercle indicateurs*, in which the Cantons, cities and federal agencies are not legally obliged to participate, the collegial benchmarking approach — to come back to Fenna's typology — has proven to be the best. That the long-time partners remain committed and that new ones are always joining can only be explained by the fact that they can articulate their needs in this common project; that useful products are generated for them; and that they do not have to fear any sanctions or backlash. Collegial benchmarking is not to be recommended as the most suitable solution for all conceivable benchmarking applications. When benchmarking is to be used in jurisdictions where the central government, for example, justifiably exercises some control because of the flow of money to local authorities, the principle of purely voluntary cooperation will not suffice. Yet, in this case too, it is recommended that the support of the participating regional authorities be guaranteed by giving them the opportunity to participate as much as possible, for instance, in setting the rules.

With appropriate rules, the incentives and learning benefits outweigh any possible negative effects

Indicators can trigger strategic behaviour (Fenna, this volume), especially when the indicator system is linked to sanctions, or when cash flow is impacted. Neither is, as is well-known, the case with the *Cercle indicateurs*. Therefore, it is not entirely surprising that here — thanks to the approach of the partners (collegial benchmarking) — the mutually beneficial incentives and learning benefits prevail. Within the *Cercle indicateurs*, for example, this was the case with the reporting on sustainability by the local authorities and their inclusion in managing policies. In 2005, the canton of Aargau was one of the first to prepare a report on sustainability — and this as part of its New Public Management policy. Since then, seven other Cantons mentioned above, as well as two cities, have been inspired by the positive rivalry to develop similar approaches.

The stimulating and coordinating role of the federal government is both desired and welcomed

In Switzerland, there are always some tensions in the relationship between the federal government, the Cantons and the cities. Time and again, the federal government is refused the right to negotiate in the absence of any explicit legal grounds. In the case of the *Cercle indicateurs*, the Cantons and cities have evidently always appreciated the involvement of the federal government and its coordinating and supporting role. It was very clear to all that a purely autonomous organisation consisting of Cantons and cities would have hardly been capable of uniting all stakeholders under one indicator system. A precondition for the appreciation of the federal government's involvement was, however, the mutual commitment to collegial benchmarking.

Methodological quality legitimises

As mentioned earlier, the *Cercle indicateurs* was formed by several bottom-up initiatives that were created with a lot of enthusiasm but with few resources and often-inadequate methodological and technical support. This initially exposed the *Cercle indicateurs* to a variety of criticisms, for example, by the statistical agencies, and proved to be an obstacle in terms of recruiting new participants. The institutionalised cooperation with the Federal Statistical Office proved to be very useful as much for the factual improvement of the quality of the indicator system as for its legitimisation. That the group of participants has grown even more recently and includes almost all the Cantons must also have something to do with this support from official statistics. That other benchmarking systems must also work

with official statistics should not be immediately assumed. But the investment in a clean and unquestionable methodology is strongly recommended.

Selection of 2nd, 3rd best Indicators as a concession to the reality of data availability

All indicator systems and, in particular, those used in Switzerland's federal system, must balance conceptual demand with the reality of the highly limited availability of data. This is why the *Cercle indicateurs* decided to leave gaps in some target areas, while in other cases it had to settle for a compromise. It was exactly because of these critical considerations that the participation of the Federal Statistical Office was very helpful. An open and transparent discussion of the criteria and the selection of indicators is recommended for all indicator initiatives.

Aggregated benchmarking between methodological scepticism and communications use

The primary target group of the *Cercle indicateurs* is not the general public, but rather the administrative experts and politicians of the Cantons and the cities. Yet, the legitimate need arose, based on the indicators, to raise awareness among members of the public and the media as well. Rankings are a useful instrument to do this. The method used to aggregate the indicators of the *Cercle indicateurs* to a total value can be questioned. There is a certain conflict between the objectives of communication and methodological soundness. In the case of the *Cercle indicateurs*, there is an open and honest approach to aggregation, and participation in aggregated benchmarking is voluntary. It would also be conceivable to have an independent office prepare the aggregated benchmarking and assign 'marks' without the input of those participating. In the interest of permanently maintaining the support of the participants and their impartial debate about the results, the experience has been good using the participative approach. Also even if it is voluntary, there is usually pressure to take part, since those that stand on the sidelines must explain themselves to their citizens or the local media.

Indicators can stimulate political debate

Literature on the topic disputes if and when knowledge-based approaches influence policy (Steurer and Trattnigg 2010). We saw with the *Cercle indicateurs* that, for the Cantons and the cities that had not yet gone through the sustainability process, surveying indicators can represent a step towards a sustainability policy. The results of the survey lead to a discussion of the values and ways to improve them. The

creation of an indicator system can, moreover, clearly lead to defining and structuring the topic of sustainability (Wachter 2010, p. 203). Several Cantons and cities use, as shown above, the data from the surveys for their own reports on sustainability. These reports are prepared, in part, in time with the legislative periods and serve as the basic data within the framework of New Public Management. The inclusion of the data of the *Cercle indicateurs* in the cantonal policy administration process enhances the value of the indicator system. For this effect to trigger action, rules need to be set to encourage an unbiased discussion among the participants and to provide the opportunity to exchange experience gained and to learn.

6.7 Conclusion

Given the substantial degree of autonomy enjoyed by the Cantons in Swiss federalism, it is not surprising that this example of inter-jurisdictional benchmarking has developed in a very bottom-up and collaborative fashion. That it has occurred at all is due in part to the sustainability mandate given to both levels of government by the Constitution. The federal government's role has been an important, but an entirely un-coercive one — confined to facilitating the comparability of data and welcomed on that basis by the participants. The Federal Office for Spatial Development plays precisely the role Fenna (this volume) describes as that of a benchmarking 'node'.

Cercle Indicateurs' strengths lie precisely in its collegial and voluntary nature. Not surprisingly, then, its chief audience is not the public, but the participating organisations themselves. That said, the results are made public and do get employed to draw potentially invidious performance comparisons.

Like other instances of successful benchmarking, the *Cercle Indicateurs* project has also enjoyed the opportunity for iterative improvement and confidence-building which, as Fenna (this volume) notes, is an important element in successful benchmarking. Like other instances of benchmarking as well, the exercise has been constrained by data availability but that is a situation that can be and has been addressed thanks to the project's iterative nature.

References

Cercle indicateurs 2005, *Kernindikatoren der nachhaltigen Entwicklung von Kantonen und Städten: Bericht des Cercle indicateurs*, Bern.

-
- Eurostat 2005, *European Statistics Code of Practice for the National and Community Statistical Authorities*, Eurostat, Luxembourg.
- Federal Office for Spatial Development 2008, 'Sustainability Assessment: Guidelines for Federal Agencies and Other Interested Parties', Federal Office for Spatial Development, Bern.
- Federal Statistical Office and Konferenz der Regionalen Statistischen Ämter der Schweiz Korstat 2002, *Charta der öffentlichen Statistik der Schweiz*, Federal Statistical Office, Neuchâtel.
- Federal Statistical Office, Federal Office for the Environment and Federal Office for Spatial Development 2003a, *MONET: Monitoring Sustainable Development in Switzerland — Final Report: Methods and Results*, Federal Statistical Office, Neuchâtel.
- 2003b, *Sustainable Development in Switzerland: Indicators and Comments*, Federal Statistical Office, Neuchâtel.
- Interface 2010, *Messen, Werten, Steuern: Indikatoren – Entstehung und Nutzung in der Politik*, TA-SWISS Center for Technology Assessment, Bern.
- Linder, W. 2010, *Swiss Democracy: Possible Solutions to Conflict in Multicultural Societies*, Palgrave Macmillan, Basingstoke.
- Steurer, R. & Trattnigg, R. (eds) 2010, *Nachhaltigkeit Regieren: eine Bilanz zu Gouvernanz-Prinzipien und –Praktiken*, Oekom, München.
- Wachter, D. 2010, 'Politischer Nutzen von Evaluationen und Monitoring am Beispiel der Schweizer Nachhaltigkeitsstrategie', in Steurer & Trattnigg (eds), *Nachhaltigkeit Regieren: eine Bilanz zu Gouvernanz-Prinzipien und –Praktiken*, Oekom, München.
- 2007, 'Monitoring the Sustainability Strategy in Switzerland', in OECD (ed.), *Institutionalising Sustainable Development*, Organisation for Economic Co-operation and Development, Paris.

7 Benchmarking social Europe a decade on: demystifying the OMC's learning tools*

Bart Vanhercke¹
European Social Observatory (www.ose.be)

Peter Lelie²
Belgian Federal Public Service (FPS) Social Security

7.1 Introduction: puzzle, scope and limitations

The 'Open Method of Coordination' (OMC) was formally launched by the Lisbon European Council in 2000 as a new regulatory instrument for the EU. It raised high hopes as a mechanism for coordinating (sensitively) domestic policies in a wide range of areas where the EU has limited or no formal authority. In the social field, the open method 'launched a mutual feedback process of planning, examination, comparison and adjustment of the social policies of Member States, and all of this on the basis of common objectives' (Vandenbroucke 2001, p. 2). In other words, in terms of governance, the open method of co-ordination is a 'soft' tool: there is in

* The authors wish to thank the participants of the Joint Roundtable of the Forum of Federations and the Productivity Commission on 'Benchmarking in Federal Systems: Australian and International Experiences' held in Melbourne, 19-21 October 2010, for their perceptive suggestions. Special thanks go to Alan Fenna, Felix Knüpling for their constructively-critical comments; to Borja Arrue Astrain, Eva Zemandl and Cécile Barbier for their invaluable contribution to section 7.2, The institutional setup and the EU's 'rules of engagement'; and to Susanna Gürocak and Marie-Andrée Roy for their detailed suggestions. The views in this chapter are the sole responsibility of the authors. Address for correspondence: vanhercke@ose.be.

¹ Bart Vanhercke is Co-Director of the European Social Observatory (OSE) as well as associate academic staff at the Centre for Sociological Research (CESO), the University of Leuven. He is former European advisor to the Belgian Minister of Social Affairs.

² Peter Lelie is an adviser in the Belgian Federal Public Service (FPS) Social Security. From 2006 until 2010 he worked as a policy officer (seconded national expert) in the European Commission in support of the Social Open Method of Coordination. This chapter reflects the views of the author and these are not necessarily those of the FPS.

principle no ‘hard’ legislation involved, but rather ‘governance by persuasion’ (Streeck, 1996, p. 80) or ‘governance by objectives’. Precisely this (in our view misunderstood) ‘softness’ of the process led to an increasingly sceptical attitude of scholars and politicians towards the OMC: it is often seen as a weak instrument and its flaws supposedly contributed to the failure of the Lisbon Strategy.

Against such a background, it came as a surprise to quite a few observers that the Social Affairs Ministers of the Member States of the EU boldly declared, in June 2011, that the Open Method of Coordination for Social Protection and Social Inclusion (Social OMC) had proved a flexible, successful and effective instrument and that it would be reinvigorated (read: relaunched) in the context of the new Europe 2020 Strategy (Council of the EU 2011). The political objectives of the Social OMC (Council of the EU 2006) were reconfirmed: the method will continue aiming at a decisive impact on the eradication of poverty and social exclusion; the promotion of adequate and sustainable pensions; and the organisation of accessible, high-quality and sustainable healthcare and long-term care in the Member States. So Ministers confirmed the wide scope of the Social OMC: it will continue to cover not only social inclusion, but equally pensions, and health and long term care.

This chapter tries to explain the political relaunch of the OMC. It argues that while Member States recognise some of the evident flaws of the OMC process, after some initial hesitation and based on first experiences under the Europe 2020 Strategy, a majority of them decided they could not afford to lose it. To be more precise: a rather broad coalition of Member States felt that in the absence of the Social OMC’s contribution in terms of analysis and consensus framing capacity, the Social Affairs Ministers would be deprived of the necessary tools to counterbalance the one-sided focus on social protection as a cost factor in the EU’s discourse throughout 2010 and 2011. The latter was indeed striking in the first policy documents that were produced under the new strategy. In the Commission’s first Annual Growth Survey, for instance, it seemed as though pensions and health care were regarded merely as a burden on government budgets and reforms intended to ‘balance the books’ (CEC 2010a, p. 6). As importantly, social policy was narrowed down to policy against poverty and social exclusion in the setup of the new strategy. In this light, we analyse exactly *what value-added* the Social OMC has had as a ‘benchmarking’ tool so that Member States wanted to continue the process, faced as they were with a European Commission that — to put it mildly — was not insisting on continuing a strong Social OMC. By providing a description of what OMC benchmarking looks like in practice, we also supply a corrective to much of the existing literature on the topic, which is too often based on out-dated and incomplete assumptions about the OMC’s learning tools.

The aim of this chapter is to provide a picture of a range of benchmarking tools and the way a variety of EU and domestic actors are involved in them. For reasons of space, we do not provide a detailed discussion of the development of these tools (for example, indicator development). The chapter also focuses largely on the EU-level and it only considers one possible explanation for the OMC's recent re-launch: its capacity to counterbalance a one-sided Europe 2020 Strategy through its benchmarking potential. We thereby omit other possible explanations, including the interest of some actors to promote 'soft' governance with a view to avoiding more stringent EU involvement in social policy.

7.2 Setting the scene: from an unidentified political object to the OMC

The European Union was once described as an 'Unidentified Political Object' by former European Commission president Jacques Delors.³ The fact is that it is a highly original political and economic entity, different from any other historical experiences and existing federal systems. This section tries to understand the main features of this object and thereby provides the context for a description of the emergence and further development of the (Social) OMC.

The institutional setup and the EU's 'rules of engagement'

The European Union is a supranational entity composed of now 27 independent Member States. It is often conflated with a federal state system, but two key factors distinguish it: it is merely an 'emergent' federation; and its powers depend entirely on the willingness of Member States to concede sovereignty in policy areas more traditionally held by the nation-state (Fenna, this volume; Majone 2006; Laursen 2011). With each successive treaty, the European Union undergoes steady *communitarisation* ('supranationalisation'), increasing its competence and involvement in more policy areas. The European Union is, therefore, a rapidly changing institutional arrangement that poses a challenge for characterisation.

Currently, there are two essential treaties constituting the primary legislation of the EU: (1) the *Treaty on European Union* (TEU), laying out the institutional architecture and the general principles of the Union; and (2) the extensive *Treaty on the Functioning of the European Union* (TFEU 2010), which specifies its competences and function. The combination of these treaties is more commonly

³ Jacques Delors coined this famous expression on the 9th of September 1985 in his inaugural speech to the Intergovernmental Conference of Luxembourg (Bulletin EC no. 9, 1985, p. 8).

referred to as the ‘Lisbon Treaty’, which came into force in December 2009. Although it has reinforced its competences, the Union remains much weaker than the central governments of traditional federal states. It is ‘more than an international organisation or confederation of states, without having become a federal entity’ (Börzel, 2003, p. 2).

Institutional framework: formal and informal ‘rules of engagement’

The main institutions of the European Union — which to some extent reflect the separation of powers common to modern democracies — are the European Commission; the European Parliament; the European Council; the Council of the European Union; and the European Court of Justice. As for the European Central Bank, it plays an essential role since it has the full control over the Euro monetary policy.

The **European Commission** is often referred to as the ‘guardian of the treaties’. It is the main policymaking institution with the exclusive right of legislative initiative and stands for the observance of European law. It is the equivalent of the executive power, the symbol of the ‘supranationalisation’ of competences, and it also activates the legislative procedure and monitors implementation of EU legislation. The **European Parliament** (representing citizens) and the **Council of the European Union** (representing the Member States) are the co-legislative bodies. The former’s powers have been significantly reinforced under successive treaties; now, its agreement is necessary for the adoption of most European legislation.

Intergovernmental relations are formalised, first of all, in the context of the **European Council**, which consists of regular meetings between the heads of state or government that are chaired by the new ‘EU president’; its purpose is to establish general orientations on the future of the European Union. The **Council of the European Union** consists of different sectorial formations, which are composed of the relevant national Ministers. Thus, the Ministers for Economy and Finance meet in the so-called ECOFIN Council formation, and their colleagues responsible for employment, social policy, health and consumer Affairs in the ‘EPSCO’ formation. The discussions in the Council on some topics are typically prepared by EU committees — such as the Economic Policy Committee (which reports to the ECOFIN Council), or the Social Protection Committee (SPC, which supports discussions in the EPSCO Council).⁴

⁴ The Social Protection Committee (SPC) is a Treaty-based committee that supports the EPSCO Council of Ministers, next to the Employment Committee. It was set up in 2000 and is formally based on Article 160 TFEU. The SPC is composed of officials from each Member State (mainly

Subsidiarity and semi-sovereign welfare states

The EU lies somewhere between a federal system and an international organisation, with the Member States retaining their competence over many sensitive policies, such as military and defence, education and social welfare. They continue to manage all the policies that have not been expressly transferred to the European Union. In addition, while they are legally constrained in many areas by the European regulations, the Member States have the necessary financial and political instruments to implement their own policies independently. The Union relies on national support to legislate,⁵ even in the fields that are part of its ‘exclusive competences’ (Börzel 2003, p. 3).

The principle of ‘subsidiarity’ — which establishes that the Union should not exercise any responsibilities that are already satisfactorily fulfilled by the constituent units — has become one of the most important in the Union, and it ‘explicitly discourage[s] the expansion of the European Union into certain new areas’ (Moravcsik 2001, p. 172). This principle is complemented by the principle of ‘proportionality’ which seeks to set actions taken by the institutions of the Union within specified bounds (that is, the involvement of the institutions must be limited to what is necessary to achieve the objectives of the Treaties).

Title X of the *Treaty on the Functioning of the EU* provides the legal framework for the European Union’s competence in social policy, which is at the heart of the current chapter. Article 153 specifies the objectives to be achieved in this area by the ‘cooperation between Member States ... excluding any harmonisation of [their] laws and regulations’. In fact, the European institutions may produce social legislation but only to adopt ‘minimum requirements for gradual implementation.’ Moreover, the unanimity of the Council is required in order to legislate over some of the issues set out in the article.

In spite of these constraints, the European Union has an increasing impact on the social policies of its Member States through a variety of instruments that will be discussed in the next section. The result of such EU involvement is that Member States are at most semi-sovereign with regard to the development of their welfare systems (Leibfried and Pierson 1995). This finding has become even more evident in the course of 2011, when EU pressure on the national reform process increased considerably through strengthened fiscal surveillance under the Stability and

from the national Employment and Social Affairs Ministries) and from the Commission. It has a leading role in the Social OMC.

⁵ The ordinary legislative procedure (formerly ‘codecision procedure’) is the most usual legislative procedure in the European Union. It requires the approval of both the European Parliament and the Council of the European Union.

Growth Pact as well as the Euro Plus Pact (with its focus on sustainability of pensions, health care and social benefits).

The EU's toolbox in social policy

Broadly speaking, the European Union currently has four instruments at its disposal to produce an action in the social field directed towards the Member States: classic European law (directives and regulations); European social dialogue (both at the inter-professional and sectoral level); financial instruments (in particular the European Social Fund, ESF); and cooperation, including (in its strongest form) the Open Method of Coordination. These instruments are different in nature, because of their distinct historical origins. Some date from the beginning of the EU (classic European law and the ESF), other from the 1990s (collective agreements and OMC). They also differ in their legal scope. Some are binding (classical law and collective agreements) while others are inciting (OMC and ESF); some are distributive (ESF), while others are regulatory (the other three). They also differ in terms of the actor-networks that are connected to them, both officially and in practice.

As far as **EU legislation** in the field of social protection and social inclusion is concerned, the possibilities for the European Union to take action are severely constrained by institutional (see the section on *Subsidiarity and semi-sovereign welfare states*) and political hurdles. Member States and EU institutions have not been willing or able — on account of national interests, political sensitivities and the huge diversity of social protection and social inclusion systems — to legislate in these areas. The EU's legal competencies in this domain therefore constitute an unfulfilled potential. This said, the absence of 'pure' EU social security or social inclusion legislation does not mean that there is no process of 'positive integration' in these fields. Indeed, other dimensions of European integration have an indirect impact on social security and social inclusion policies, but these norms have other and prevailing objectives. One thinks first of rules of coordination of social security schemes, directives on gender equality, the fact that principles of free movement have been applied to healthcare services and the existence of legally binding acts at a procedural level' (for a more detailed account, see Reyniers *et al.* 2010). Secondly, there is the **European social dialogue**. Following the Maastricht Treaty, the *European interprofessional social dialogue* takes place between representatives of European social partners. Over the years several framework agreements have been negotiated and transposed into directives, on parental leave and temporary work for example. Other 'autonomous' framework agreements (for example, telework) are meant to be implemented by the social partners following their own national procedures. The European *sectoral social dialogue* received a major boost

in 1998 with the creation of so-called European sectoral social dialogue committees, which between them have produced more than 500 joint texts, including framework agreements.⁶

Thirdly, Member States draw financial resources from the **EU's financial tools**. The main instruments are the European Regional Development Fund (ERDF); the Cohesion Fund; and, especially relevant in the context of this chapter, the European Social Fund (ESF). Starting as a retroactive instrument used by Member States to finance their vocational training programs in the 1950s and 1960s, the ESF evolved into a more pro-active instrument that supports a rather broad set of social (inclusion) and labour market policies. In the current EU governance setting, the ESF's principal objective is to provide financial support for actions taken within the framework of the European Employment Strategy.⁷ While the impact of these financial instruments has been significant for the least developed regions and countries, their impact has also been noticeable in the most developed countries (Verschraegen *et al.* 2011).

The fourth and final EU instrument is **cooperation** between the Member States, which has been around for a long time. And yet, based on the perceived need to work together around common social policy challenges, advanced forms of cooperation aimed at *coordination* and *convergence* of domestic policies have emerged over the years: notably the open method of coordination. The OMC is an example of 'New Governance', which involves 'a shift in emphasis away from command-and-control in favour of 'regulatory' approaches, which are less rigid, less prescriptive, less committed to uniform approaches, and less hierarchical in nature'. The idea of new (or experimental, or soft) governance 'places considerable emphasis upon the accommodation and promotion of diversity, on the importance of provisionality and reversibility ... and on the goal of policy learning' (de Búrca and Scott 2006). We turn to the OMC in the next section.

OMC: Emergence, general features, variation and inflation

Defining the elephant

In March 2000, in Lisbon, the EU Heads of State and Government set 'a new strategic goal' for the Union: to become, within a decade, 'the most competitive and dynamic knowledge-based economy in the world capable of sustainable economic

⁶ See <http://www.etui.org/Topics/Social-dialogue-collective-bargaining/Social-dialogue>.

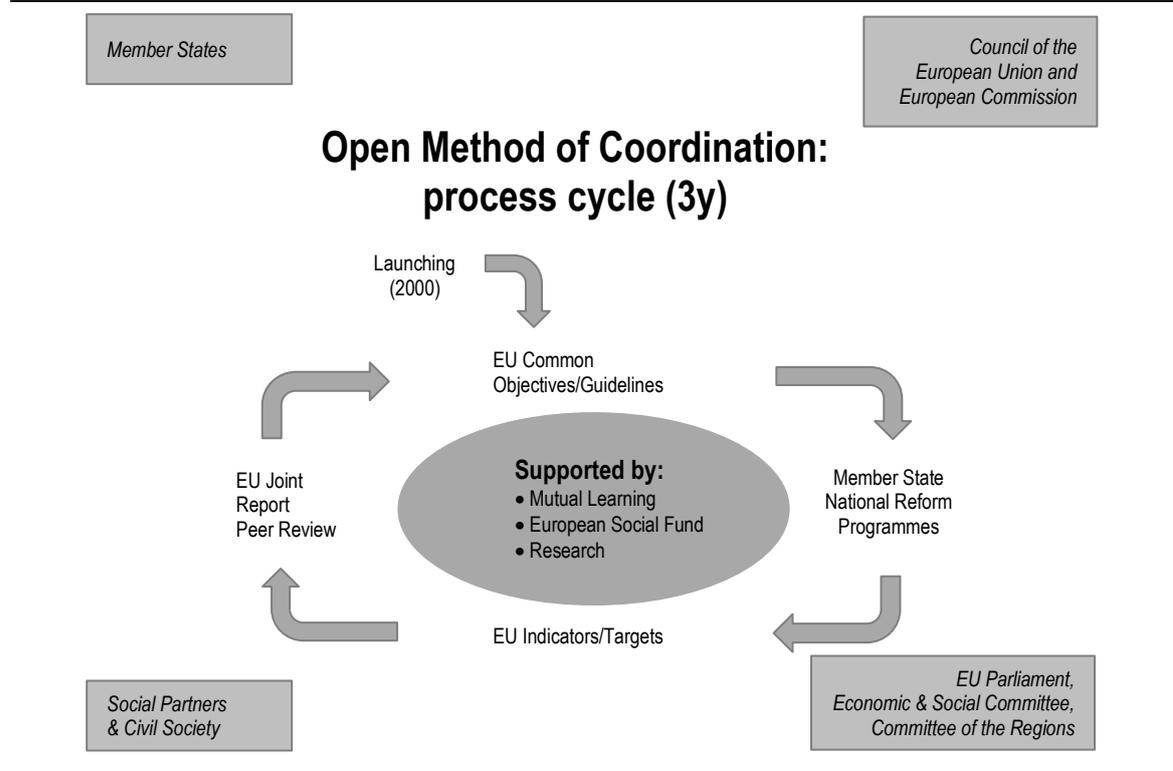
⁷ The ESF regulation of 1999 stipulates that the Fund should support policy measures of the Member States that are in line with the EES (EP and Council of the EU, 1999).

growth with more and better jobs and greater social cohesion’ (European Council 2000, para. 5). This strategy was to be implemented by improving the existing processes, ‘introducing a new open method of coordination’ (European Council 2000, para. 7).

What is this method, then? Since there is no legal definition of the OMC in the Treaty or other binding texts, reference needs to be made to the Presidency Conclusions of this Lisbon Summit. This text refers to the OMC as ‘the means of spreading best practice and achieving greater convergence towards the main EU goals’. Still according to the same source, this involves: fixing *guidelines* (with specific timetables); establishing quantitative and qualitative *indicators* and *benchmarks* (against the best in the world), national and regional *targets*; and periodic *monitoring, evaluation* and *peer review* organised as *mutual learning processes* (European Council 2000, para. 37). Ultimately, the purpose is that Member States learn from one another, and thereby *improve domestic policies*. Conceptually, the OMC finds its roots in the Broad Economic Policy Guidelines which were introduced by the Treaty of Maastricht (1992) and involved non-binding recommendations to monitor the consistency of national economic policies with those of the European Monetary Union. Other examples of OMC *avant la lettre* (that is, before the Lisbon European Council labelled the policy instrument as such) include the European Employment Strategy (the ‘Luxembourg process’); the Cardiff Process for structural economic reforms; the Bologna Process for cooperation in European higher education; and the code of conduct against harmful tax competition (Zeitlin 2005, p. 20).

Figure 7.1 highlights the different ‘components of what an ‘ideal’ OMC looked like in the past decade. The arrows make clear that OMC is a cyclical process (typically three years) where mutually agreed Common Objectives (political priorities) are defined, after which peer review (discussion among equals) takes place between the Member States on the basis of national reports (called National Strategic Reports, National Reform Programmes, etc.). Soft ‘recommendations’ (issued by the Commission and the Council) and comparable and commonly agreed indicators (and sometimes quantified targets) enable to assess progress towards the Objectives.

Figure 7.1 **OMC Cycle under the Lisbon strategy (2000–2010)**



After Lisbon: proliferation

The 2000 Lisbon Council Conclusions stipulated the introduction of the OMC ‘at all levels’ (European Council 2000, para. 7), and explicitly referred to the use of the OMC with regard to social exclusion, information society/e-Europe (para. 8), innovation and research and development (para. 13). Furthermore, even though the term ‘OMC’ was not explicitly mentioned with regard to social protection (pensions more particularly), enterprise promotion, economic reform and education and training, the wording of the Lisbon Council Conclusions was such that it gave, *de facto*, authorisation to launch or at least strong political backup to continue open co-ordination in a host of policy areas. As a result, the OMC is now up and running in more than 10 policy areas.

This proliferation of soft law tools did not really come as a great surprise. In view of the enlargement of the EU in 2004 and 2007, with further enlargements in preparation,⁸ very few EU initiatives are taken with a view to finding agreement on EU social legislation, and it seems highly unlikely that legislation in the social sphere will increase substantially in the next few years. Therefore some predicted, at the turn of the century, that ‘policy co-ordination and benchmarking [would be] a

⁸ For example towards the Balkan countries.

typical mode in future EU policy-making, as an alternative to the formal reassignment of policy powers from national to EU level' (Wallace and Wallace 2001, p. 33). Indeed, for a number of politically sensitive areas (where no unanimity could be found on EU legislative initiatives), decision makers agreed that 'doing nothing' at EU level is not an option either — if only because EU Member States are faced with a number of common challenges. Such is, for example, the case for social inclusion, pensions and health and long term care, where the OMC is providing a Europe-wide approach. The next section discusses how this approach has been developed in practice over the past decade.

7.3 Benchmarking within the Social Open Method of Coordination (2000–2010): how did it really work?

For many authors 'policy learning' and 'benchmarking' are some of the core features, or even the *raison d'être*, of the European process. In this light it seems all the more striking that oftentimes students of the OMC remain rather vague about *what they are actually talking about*, when they make claims about the success or failure of 'benchmarking' in this context. This section fills that gap in our understanding by describing *how benchmarking is actually done* in the Open Method of Coordination on Social Protection and Social Inclusion (Social OMC), mainly illustrating the process with reference to the social inclusion strand.⁹ The focus is essentially on the post-2005 and pre-2011 Social OMC. Before 2005 three separate social OMC's co-existed: one on Social Inclusion (2000); one on Pensions (2002); and one on Health and Long Term Care (2004). In 2006, the Open Method of Coordination on Social Protection and Social Inclusion was established, regrouping and integrating the three processes. While the Social OMC is to continue under the Europe 2020 strategy, it is already clear that there will be important changes. We will briefly comment on these in section 7.4.

The Member States and the European Commission engage in what Alan Fenna (this volume) calls 'collegial benchmarking'. They do so through the following six steps:

- Agreeing on the framework (the mandate) — the Common Objectives
- Selecting key issues in a multidimensional policy domain

⁹ In fact the Social OMC is embedded in a broader setting where several similar processes are ongoing that can be labelled as 'social' and that have close links with the Social OMC: these include the employment, anti-discrimination and gender equality processes. Social benchmarking also takes place in the context of the structural funds (for example, the European Social Fund).

-
- Building the knowledge base — defining the issues and developing common indicators for quantitative benchmarking
 - Supporting the process through non-governmental expert and EU (civil society) stakeholder networks
 - Engaging in the benchmarking process — (different types of) peer reviews and OMC ‘projects’
 - Drawing conclusions — joint reports on social protection and social inclusion and Commission ‘recommendations’ (lessons learned).

We discuss these steps in turn and will argue that all need to be considered together if one wants to understand the ‘learning potential’ and the impact of this process.

Agreeing on the framework (the mandate): common objectives

The starting point of benchmarking in the context of the Social OMC is the acceptance by the Member States of a set of ‘common objectives’. These provide the mandate, and thereby define the framework for the exercise. In total four subsets of common objectives have been defined: there are so-called *overarching* objectives, and one set for each of the three strands of the Social OMC: social inclusion, pensions, and health and long term care (Council of the EU 2006). The objectives concern social protection and inclusion *outcomes* as well as the way in which social protection and inclusion policy is developed: principles of *good governance*.

Because the common objectives are meant to be fairly constant over time and agreed-to unanimously (see the section on *Subsidiarity and semi-sovereign welfare states*) between governments from all Member States from across the political spectrum, they are quite general in nature. This aspect has been heavily criticised by some, while others (for example, Greer and Vanhercke 2010) have pointed out the fact that it is precisely because of the ‘general’ or ‘vague’ nature of the objectives that they can be the starting point for what can be dubbed a European ‘consensus framing exercise’. A major value-added of the objectives is that they structure the policy field in such a way that a balanced policy approach to the different strands is promoted and has become widely accepted among EU and domestic policymakers alike. For example, good health care presupposes *accessible* and *financially sustainable* health care systems providing *high levels of quality*.

Selecting key issues in a multidimensional policy domain

‘Social protection and social inclusion policy’ has been defined, from the outset, as a broad multidimensional policy domain. Thus many issues and policies are potentially relevant with a view to a benchmarking exercise. The first priority has therefore been to work on a consensus regarding the main challenges. As far as social inclusion is concerned, Member States have singled out key issues such as: how to bring about the social inclusion of people on active age far from the labour market (active inclusion); how to tackle child poverty; and how to fight housing exclusion and homelessness. As far as pensions are concerned, one of the key issues is how to ensure both adequacy and financial sustainability of pension systems in the long run. In the area of health there has been a focus on tackling health inequalities. As the awareness of the so-called implementation gap (the perceived lack of results of the Social OMC) grew, attempts have been made to increasingly focus the process on specific issues, especially so in the area of social inclusion. Benchmarking will indeed be most effective if it is concentrated on a limited number of priority issues.

Building the knowledge base: defining the issues and developing common indicators

Once an agreement on common objectives and priority issues has been reached, the knowledge base needs to be developed. Any credible benchmarking in this context will need to build on a consensus with regard to the *definition* of the social protection and inclusion challenges and with a reliable description of how they present themselves across the European Union. This implies the development of indicators or quantification. There is indeed some truth in the idea that if at EU level a policy challenge cannot be quantified it does not exist. Or, as the Open Society Institute famously put it: ‘no Data, no Progress’ (OSI 2010). So, to the extent possible, challenges and policy outcomes need to be quantified, even if reducing and simplifying disparate information into numbers (commensuration) involves some clear risks and the social process behind it should be acknowledged (Espeland and Stevens 1998). Hence, the development of commonly agreed indicators has always been at the heart of the Social OMC. The commonly agreed indicators are the yardsticks for measuring policy challenges and policy success and failure. Without them, the benchmarking exercise would be reduced to endless discussions about concepts, reliability of data, etc. The indicator set of the Social OMC has been developed by a dedicated working group: the Indicators Subgroup (ISG) of the Social Protection Committee (see the section on *Institutional framework: formal and informal ‘rules of engagement’*), in which all Member States and the Commission participate.

The Indicators Sub-Group proceeds on the basis of a set of methodological principles. These frame the quality criteria an indicator (and an indicator set) needs to comply with. In order to ensure international comparability, harmonised EU data sources such as EU-SILC¹⁰ (previously ECHP) have been developed. The structure of the set follows the structure of the common objectives. As a result, there are four subsets: one overarching set and one set for each strand of the Social OMC (CEC 2009a). At the request of the European Commission, the Herman Deleeck Centre for Social Policy has developed a *Vade Mecum* of the Social OMC indicator set.¹¹ Data are available on the Eurostat website¹² and on the website of DG Employment, Social Affairs and Inclusion.¹³

The commonly agreed indicators have at times been criticised in academic circles for not being sufficiently sophisticated or innovative. This, however, somehow misses the point. The main value-added of the set is that Member States *have reached a consensus* on them. This is not self-evident in view of the different ways in which social protection and inclusion challenges and policies present themselves across the EU. Debates with regard to indicator development in the ISG have been characterised by a peculiar mixture of technical arguments and underlying political motives.

Over the years several developments have taken place that are of interest from the perspective of benchmarking and which we discuss in turn before briefly describing how the indicators are used for quantitative benchmarking.

Developing common indicators: key trends

In the framework of the Social OMC, policy documents have often stressed the *multidimensional nature* of social protection and social inclusion. It has, however, not been possible to develop indicators covering all dimensions at once. The agreement on the original set of social inclusion indicators in 2001 (the Laeken indicators) was a major breakthrough, but it necessarily had to focus on the ‘low hanging fruit’: indicators for which an agreement was within reach and for which harmonised data sources were available. As a result, the social inclusion indicators

¹⁰ EU SILC refers to the European Union Statistics on Income and Living Conditions. See: http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc.

¹¹ See: <http://www.ua.ac.be/main.aspx?c=.VADEMECUM&n=79891>.

¹² See: http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_social_policy_equality/omc_social_inclusion_and_social_protection.

¹³ <http://ec.europa.eu/social/main.jsp?catId=756&langId=en> See related documents. The Commission website has been widely criticised (including by the SPC) for being insufficiently user friendly.

set has been criticised for being too income and labour market focussed. For a long time, dimensions like education, literacy, health and housing were indeed underdeveloped or invisible. Progressively and not without problems the set has been expanded in these other areas. Over the years it has become increasingly clear that international comparability is a big challenge.¹⁴ Agreeing on new indicators takes time and an important lead time is needed to develop new harmonised data sources.

The development of the indicators has been influenced by the progressive *extension of EU Membership*. The original set of social inclusion indicators was developed in the context of the EU15 (2001). In 2004, ten new Member States joined the EU and two more countries followed in 2007. Needless to say that adding such a substantial number of countries characterised by varying levels of economic development and income inequality had an impact on the process. On a number of the indicators the EU average shifted markedly and some indicators and basic concepts had to be reconsidered. One evident example is the concept of poverty itself. In the original set of Laeken indicators, the headline indicator for measuring poverty was the relative at-risk-of-poverty rate. Since some of the new Member States were characterised by low income inequality they were among the best performers measured using this indicator, while their general standard of living (and the absolute level of the relative poverty threshold) was actually quite low. There was a lot of pressure to develop a more balanced picture of poverty that would show both its relative and absolute dimension. Consequently, in 2009 an agreement was reached on a material deprivation indicator, which reflects the share of persons who have living conditions severely constrained by a lack of resources.

Originally, the set of social inclusion indicators was exclusively *outcome oriented*. However, especially after the midterm review of the Lisbon strategy in 2005 and the discussion about the implementation gap, the SPC's Indicators Sub-Group progressively started working on input and output indicators so as to facilitate analysis of the links between policies and policy instruments and outcomes (which policies produce which outcomes?). This development has also been prompted by indicator development in the non-social inclusion strands, where there is a more natural focus on policy instruments (for example, pensions are policy tools). A 'national indicators' label was developed in cases where a *common definition* has been developed, but where harmonised data sources are missing and/or due to major institutional differences, cross country comparability is problematic.

¹⁴ One example of the difficulty in producing internationally comparable indicators concerns subjective indicators, for example in health.

In order to increase political commitment and make the general objectives more concrete, the use of *national targets (or quantified objectives)* has been promoted from the start of the Social OMC. After the Barcelona European Council in 2002, pressure on the Member States increased. In its ‘guidance notes’ for National Strategy Reports on Social Protection and Social Inclusion (see below), the European Commission has stressed the importance of targets. Annexed to several of the guidance notes was a tool for developing targets. An increasing number of countries have effectively introduced such targets in their strategies. At the start of the Social OMC, the idea was to also produce EU level targets to drive the process but it has proved to be impossible to reach an agreement in this respect before 2010. We will briefly refer to the introduction of an EU-level target on poverty and social exclusion in 2010 in section 7.4.

Indicators: genuine benchmarking in spite of sensitivities

It is not so difficult to understand that Member States are particularly nervous about being ranked on indicators, but in a way, being scored on a balanced set of indicators like the Social OMC set, rather than on one single (eventually composite) indicator, increases the acceptability for the Member States. Although some Member States clearly are excellent performers on many indicators while others combine bad scores on many indicators, it helps that, in a multidimensional approach, rankings differ. Many Member States counterbalance bad performance on one indicator with good performance on others. Also, scores change over time and Member States may move up or down the ranking.

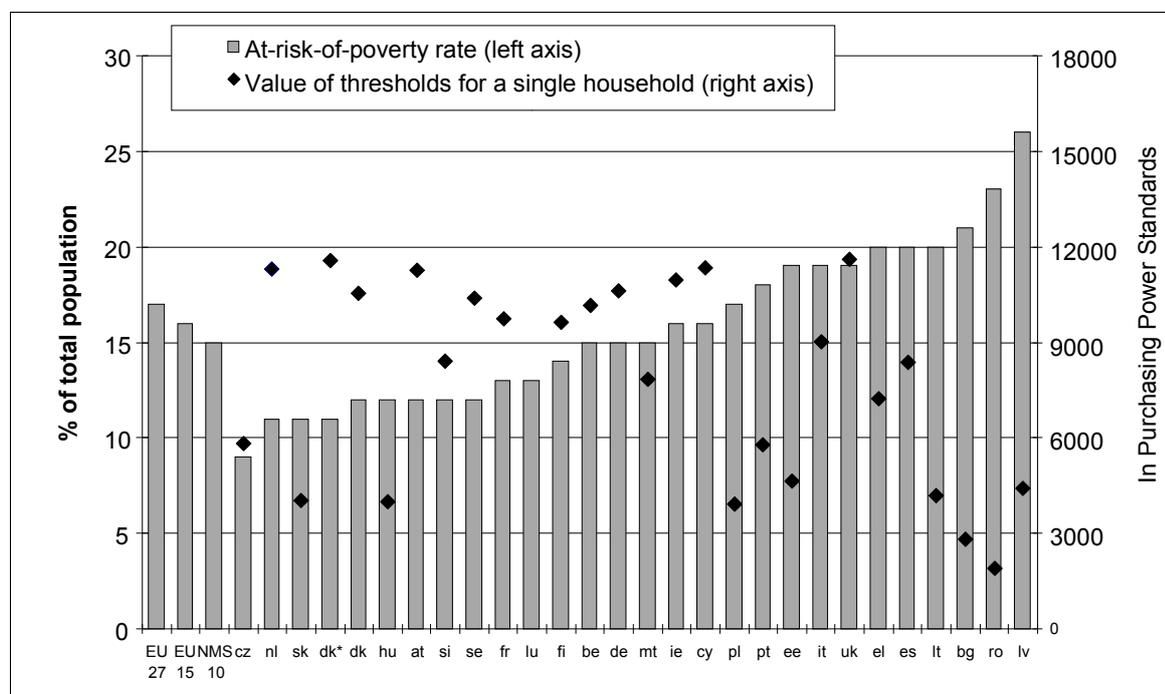
The set of indicators is continually updated and these updates of the indicators are also provided to the Member States ahead of any reporting exercise, so that they can compare their own performance with that of other Member States. Guidelines for a new round of national reports insist that the starting point for a new domestic strategy should be the assessment of progress on targets and the assessment of the social situation (there are considerable variations between Member States as to the extent in which they take on such an evidence based approach).

In spite of the sensitivities amongst (and the pressure from) the Member States, the European Commission has tried to make the best of its role of ‘independent arbiter’ in the benchmarking exercise. Thus, since the start of the Social OMC, Member States have been ranked on the basis of commonly agreed indicators in the annual joint reports, as can be seen from figure 7.2 below. These reports (which are discussed in section 7.3) always contain an analysis of the indicators.¹⁵ The most

¹⁵ See the ‘Social situation in the EU27’ section in the most recent Joint Report 2010 (CEC 2010d).

recent trends are described, Member States are compared on the different dimensions and causal factors that could explain trends are explored. In ‘full reporting’ years, the country fiches in Joint Reports (see further) also contain a statistical profile per country (selection of commonly agreed indicators). This statistical profile compares the Member States’ scores on some key indicators with the EU averages.¹⁶ Benchmarking has also been a key feature of ad hoc reports produced by the Social Protection Committee, including the successive ‘Assessments of the social impact of the Crisis’ (see, for example, SPC and CEC 2010).

Figure 7.2 Example of graph showing ranking of member states (at-risk-of-poverty-rate) in the Joint Report on Social Protection and Social Inclusion 2010



Sources: CEC 2010d (Supporting document p. 24), data EU-SILC 2008.

Important progress in the use of indicators has been made at the occasion of the thematic reporting years. A key development was the child poverty report that was produced by an SPC-ISG task force on child poverty (SPC 2008). It provided an example of how far the analysis (benchmarking exercise) could be pushed using the common indicators. The child poverty report contained an explanatory analysis in that it not only showed a ranking of Member States regarding child poverty *outcomes*, it also examined main *causal factors*: joblessness, in work poverty and

¹⁶ See: <http://ec.europa.eu/social/main.jsp?catId=757&langId=en>.

ineffective social transfers. A six level scale is used, ranging from ‘+++’ (countries with the best performance) to ‘- - -’ (countries with the worst performance).

Note that the foreword to the child poverty report, signed by the SPC Chairperson and the Commission Director-General for employment and social affairs, indicates that ‘Indicators have not been used to name and shame but to group countries according to the common challenges they face’. This is a clear sign that ranking countries, although part and parcel of the Social OMC, remains a sensitive issue.

In case Member States take a firm stand on figures that are to be published in highly visible political documents, the Commission will occasionally be flexible and allow them to have a footnote with a table or graph indicating concerns about reliability of data or pointing to certain particularities crucial for interpreting the data for a Member State.¹⁷

Supporting the process through non-governmental expert and EU (civil society) stakeholder networks

The key actors participating in the Social OMC are the Member States (through the Council and its preparatory Committees) and the European Commission. However, benchmarking in the framework of the Social OMC is also supported by the involvement of two additional types of actors which tend to be overlooked by OMC critics in this context: independent, or rather non-governmental expert networks on the one hand; and civil society stakeholder networks on the other.

Non-governmental expert networks

Benchmarking needs an independent arbiter with sufficient expertise and authority (Fenna, this volume). This role is played by the European Commission. Specifically with regard to the social policy area, the Commission’s DG Employment, Social Affairs and Inclusion has country desks that follow up developments in a specific Member State and thematic units that cover a specific issue across all Member States (‘vertical’ and ‘horizontal’ units). Other DG’s of the European Commission are less frequently involved in the Social OMC.

¹⁷ By way of illustration, Denmark insists that the at-risk-of-poverty rates should be calculated including imputed rent (income derived from home ownership). In the absence of imputed rent calculations for all Member States and in view of the lack of agreement on including imputed rent in standard at-risk-of-poverty calculations, two series of results are published for Denmark. In figure 7.2 above, dk* refers to the figure including imputed rent.

In order to assist the Commission in its policy assessment work, a growing number of thematic EU non-governmental expert networks have been put in place. The experts in these networks can be considered as the Commission's 'eyes and ears' in the Member States. Members of the networks are to provide independent (that is, non-governmental) expertise. Specifically with regard to the Social OMC, two such networks are of prime importance: the Network of Independent Experts on Social Inclusion¹⁸ and the ASISP network that covers social protection issues (pensions, health and long term care),¹⁹ but others are also relevant.²⁰ Some of the analyses of these networks (especially assessments of Member States' reports) feed into the Commission's assessments while being kept confidential so as to increase the likelihood that the Commission will get a frank assessment. Other analyses are published. Although in some Networks Member States can comment on draft reports that will be published, it is not up to them to take decisions about their contents. The Commission sees the expert networks (which meet regularly) as partners, sets their agenda and regularly briefs them on new developments.

EU civil society stakeholder networks

It has often been stressed in the literature that the OMC actually works through the involvement of non-governmental domestic stakeholders. The Common objectives have always requested that Member States involve relevant stakeholders at the national level throughout the policy cycle.

It has also been considered important that stakeholders participate in the Social OMC process at EU level. In order to give stakeholders a voice, core funding is being provided for EU stakeholder networks through PROGRESS (PROGRamme for Employment and Social Solidarity).²¹ These networks are intended to bring stakeholders together at national level and to support and organise their engagement with the Social OMC process at national and EU level. They are expected to assess input provided by governments, and to participate in consultations and benchmarking activities. In 2010 the Commission was funding twelve such

¹⁸ See: <http://www.peer-review-social-inclusion.eu/network-of-independent-experts>.

¹⁹ ASISP refers to Analytical Support on the Socio-Economic Impact of Social Protection Reforms. See: <http://www.socialprotection.eu/index.html>.

²⁰ For example, the employment experts network SYSDem, see http://www.eu-employment-observatory.net/en/about/abt03_01.htm.

²¹ The PROGRESS Programme is a financial instrument supporting the development and coordination of EU policy in the areas of employment, social inclusion and social protection, working conditions, anti-discrimination and gender equality.

networks²² specifically under the social inclusion theme. These networks cover themes such as the social economy, homelessness, children-in-poverty, social services, financial inclusion, and social inclusion policy in local government.²³

Like the non-governmental expert networks (see section 7.3, *Non-governmental expert networks*), the civil society networks are considered as partners in the Social OMC process by the Commission, which regularly organises meetings with them to exchange information. Many of these civil society networks have organised benchmarking exercises and developed scorecards of their own, consulting their members on what governments are actually doing with regard to specific issues. For example: the European Federation of National Organisations Working with the Homeless (FEANTSA) organised some ten ‘shadow’ peer reviews through consultation of its members (see box 7.1 below); Caritas Europa members scored civil society involvement in the 2006–08 national social protection and inclusion strategy design process against a series of indicators (Caritas Europa, 2007); The European Anti-Poverty Network (EAPN) recently produced a scoreboard containing their members’ assessment of Member States’ National Reform Programmes 2011.²⁴

The role of the stakeholder networks is of key importance for the benchmarking exercise. Some of them are quite active in mobilising civil society as well as in lobbying the European Parliament (an institution otherwise only to a very limited extent reached by the OMC) and in linking up the EU process with other international (eg., UN or OECD) related processes. The Commission has actively been promoting the development of a ‘Social OMC community’, bringing together all these actors regularly in the context of conferences and seminars, but also in the context of Peer Reviews, to which we turn in the next section.

²² The Commission has three yearly strategic framework contracts with the networks. They are funded by the PROGRESS programme to varying degrees (principle of co-financing), which range between very modest proportion of total funding for large organisations such as CARITAS, and very significant means for networks such as EAPN or FEANTSA.

²³ Other stakeholder networks are funded under other themes: e.g. anti-discrimination like European Network Against Racism and AGE Platform Europe (older persons). A number of the networks work together in the Social Platform of European Social NGOs. See: <http://www.socialplatform.org/>.

²⁴ See: <http://www.eapn.eu/images/stories/docs/NRPs/nrp-report-final-en.pdf> pp. 72–3.

Box 7.1 FEANTSA's Shadow Peer Review on the UK Rough Sleepers Strategy

The EU-level PROGRESS Peer Review on the Rough Sleepers Strategy (London, May 2004) inspired FEANTSA to carry out its own electronic 'shadow peer review' (2004) alongside the official EU Peer Review. The methods used in the FEANTSA peer review were similar to the EU Peer Review, although it involved far more countries. The network circulated a summary of the Rough Sleepers Strategy and an e-mail questionnaire to its Members (nongovernmental organizations), which are generally important actors in implementing any homeless strategy and have extensive experience of working closely with rough sleepers. They were asked for feedback and evaluation of (1) the aims and results of the strategy; (2) the transferability of the strategy in their different national contexts. The main findings of the review were published.²⁵

Engaging in the benchmarking process: peer reviews and OMC projects

Peer reviews of national policies are organised on the basis of input provided by the Member States. Two main types can be distinguished: those organised at the level of the Social Protection Committee ('SPC-peer reviews') and those in the context of the Progress Programme ('PROGRESS Peer Reviews'). Benchmarking is also at the heart of so-called OMC projects. We discuss each of these in turn.

Peer Reviews at the level of the Social Protection Committee

These reviews are organised by the European Commission and the Social Protection Committee (typically) in Brussels on the basis of the reports that are periodically produced by the Member States in the framework of the Social OMC. Normally, after each reporting exercise an SPC peer review is organised. All Member States (SPC delegates) and the Commission participate, occasionally supported by some experts. Peer reviews at the SPC level are rather closed exercises in the sense that they are not open to other actors, such as stakeholder networks.

In the early years of the Social Inclusion OMC, Member States were required to draw up National Social Inclusion Plans (NAPIncls) every second year (full reporting year). In their plans the Member States set out how, given their particular circumstances and challenges, they intended to make progress in the direction of the common objectives. The plans were prepared on the basis of a guidance note agreed

²⁵ See: http://www.feantsa.org/files/social_inclusion/Peer%20Review/EN_PeerReview.pdf.

between the Commission and the Social Protection Committee. The guidance contained a common blueprint so that comparison and mutual learning are facilitated.

It is important to distinguish between two ‘waves’ of peer reviews. The ‘*early*’ *peer reviews* (organised before EU enlargement in 2004) on National Social Inclusion Plans, took place during plenary sessions (with all SPC delegations present, implying some 50 to 60 participants). They did not generate a lot of enthusiasm by participants, to put it mildly. In fact, a lot of the criticism in the literature about ‘OMC Peer Reviews’ seems to refer to this particular early type of reviews, which tended to lead to endless and frankly tedious meetings where the full National Social Inclusion Plans of all Member States were presented one after the other, with nearly no time for discussion.

Since 2005-06, the ‘streamlined’ Social OMC increasingly focused on key issues, allowing for more in depth discussions. In the full reporting years (2006 and 2008), Member States have been asked to present only their policies with regard to a limited number of strategic policy priorities in their National Strategy Reports on Social Protection and Social Inclusion.²⁶ In between full reporting years, reporting has been limited to one focus theme per strand (thematic years or light reporting years). For example, in the area of social inclusion there has been a thematic focus on child poverty (2007); on homelessness and housing exclusion (2009); and on the social impact of the financial and economic crisis (2010). There has also been a multi-annual focus on active inclusion (social inclusion of people of active age far from the labour market) over the years. In ‘thematic’ or ‘light’ reporting years, the Commission and the SPC draw up a thematic questionnaire and the review is done on the basis of Member States’ replies to the questionnaire as analysed by the Commission. The national replies to thematic questionnaires are (unlike the NSR’s) normally not published on the internet.

As a result, in more recent years the SPC peer reviews have become more focused with separate working groups on different themes (key topics) and have become known as ‘in depth reviews’. Some Member States introduce an issue while some others comment and ask questions. There is hardly any formal reporting but the Commission ensures that main results feed directly into the annual joint report (see section 7.3). For an illustration of such an in depth SPC review, see box 7.2.

²⁶ In the case of social inclusion, it is suggested that the NSR should cover at most four strategic priority policies. Member States are invited to include full strategies in an annex to their NSR. NSR’s are published on the Commission website.

Box 7.2 In depth SPC review on child poverty

On 3 October 2007 an in depth SPC review on child poverty took place. Policies were assessed in six working groups with four or five countries presenting and four or five countries discussing in each group (i.e. 9/10 countries participated in each working group). The working groups focused on: family resources, access to services, access to education, children in impoverished neighbourhoods, protection of children at special risk and early prevention with a focus on pre-schooling. All discussants could comment or ask questions after each presentation.

Note that while SPC peer reviews may be rather closed shops, this does not mean that the nongovernmental expert networks and the EU stakeholder networks (see section 7.3, *Supporting the process through non-governmental expert and EU (civil society) stakeholder networks*) are entirely absent from the benchmarking process. In fact, the nongovernmental expert networks support the Commission in its assessment of the national strategies and Member States' replies to thematic questionnaires. EU stakeholder networks such as EAPN, Feantsa or Caritas will produce their assessments of the national Social Protection and Social Inclusion Strategies soon after these have become available. In doing so, they attempt to influence the lessons drawn from the exercise (through the joint reports), focusing attention on their particular concerns. Assessments are widely published and sent to the Commission, Member States, the European Parliament and a wide audience of Social OMC participants.

Peer Reviews organized through the PROGRESS Program

Each year, between eight and ten peer reviews about specific issues relevant for social protection and social inclusion are organised in the Member States (host countries) in the context of PROGRESS. Although the Commission suggests some priority issues (linked to the SPC work program) this does not constrain the Member States in the choice of topics. There is real ownership of the Programme by the Member States: they present themselves as candidate host countries, they can propose any subject that is of interest to them, and the decision on which proposals are accepted is based on preferences for participation expressed by the SPC delegates (proposals that receive most support are accepted).

Two modalities of PROGRESS peer reviews can be distinguished:

- *Good practice peer reviews* — in this case a Member State suggests one of its policies that has produced (exceptionally) good results and might possibly be transferred to other Member States. It invites other Member States, experts and

stakeholders to an in depth assessment of the policy and its potential transferability across the EU.

- *Policy problem or policy reform peer reviews* — in this case a Member State confronted with an ineffective policy and contemplating a policy reform invites other Member States, stakeholders and experts to study the problem and to suggest remedies, good practice examples. Academic scholars have referred to ‘learning ahead of failure’ in this context (Hemerijck and Visser 2001).

As the ‘way of doing things’ for both modalities of the PROGRESS peer reviews is largely the same, we discuss them together. The seminars are organised by a contractor that provides facilitators (chairing of sessions and reporting) and is responsible for the general management and the preparation of the seminar.²⁷ An independent thematic expert is recruited for putting the issue in a comparative EU context and for supporting the process. In a typical peer review on average seven or eight peer countries and two EU stakeholder networks participate. Peer country delegations consist of two people: an official from a ministry or local authority and a non-governmental expert who is selected by the peer country. Stakeholder networks are normally represented by one person. They can be selected from the social inclusion networks (see section 7.3, *Supporting the process through non-governmental expert and EU (civil society) stakeholder networks*), but other EU stakeholder representative organisations are also regularly invited (for example, EU social partner organisations). As a rule, the host country member of the relevant non-governmental expert network is invited to participate as this person can provide the much-needed critical input based on expert knowledge of the issues discussed and of the host country.

For these peer reviews, a specific methodology has been developed over the years. A standard peer review will take two days. Well before the seminar, a discussion paper is prepared by the thematic expert (describing the policy under review and putting it in international comparative perspective).²⁸ The paper will conclude with the key questions for discussion. Peer countries and stakeholder networks are invited to produce comment papers before the meeting so that all participants are in principle well informed before the meeting and the discussion can start at a higher level. The morning of the first day will be an *ex cathedra* introduction to the policy under review, followed by questions and answers. The afternoon will normally consist of a site visit (how is the policy implemented in practice) if relevant, or a more practical and interactive exchange between participants. The second day will be devoted to peer country and stakeholders contributions and to the drawing of

²⁷ The current contractor is ÖSB consulting.

²⁸ Often the host country also produces a host country paper.

lessons: discussion of pros and cons, room for improvement, transferability (if so, under what conditions?).

In order to facilitate a critical assessment the Commission insists that the host country should provide monitoring and evaluation data (this is sometimes a problem because many peer reviews are about pilot projects for which no evaluation data are available at the time of the peer review) and that local stakeholders (e.g. users of the services under review or people that are the subject of the policy) are present and can be questioned by participants. EU stakeholder networks will typically also ask their member in the host country to provide them with a critical analysis.

Between 30 and 40 people typically participate in a Progress peer review meeting. The limit on the number of participants aims at increasing the chances of in-depth discussion. Participation is by invitation only. In order to compensate for the closed setting, a lot of effort goes into the reporting of the discussions. A very rich website is available that contains papers, short and synthesis report, and — quite exceptional in a European context — minutes of the meeting.²⁹ Results of PROGRESS peer reviews impact less directly on the central messages that come out of the Social OMC than SPC peer reviews. But there has been an increasing effort on the part of the Commission to disseminate lessons learned and to make sure those lessons feed into the main process.

Table 7.1 summarizes the different features of the two types of peer review ('In depth' and 'Progress Programme') practised under the Social OMC since 2006.

²⁹ See: <http://www.peer-review-social-inclusion.eu/peer-reviews>.

Table 7.1 **Comparison of the two types of peer review practiced under the Social OMC (2006-2011)**

	<i>'In depth' Reviews (SPC)</i>	<i>PROGRESS Programme' Peer Reviews</i>
Hosting and main responsibility for agenda setting	SPC (Especially Bureau and Secretariat)	One of the participating countries ('host country') ³⁰
Venue	Brussels	In the host country
Frequency	Linked to the reporting cycle (full and thematic reporting): about one each year	About eight to ten seminars each year
Management of the meeting	SPC Secretariat and SPC Bureau	Peer review manager - organizer (external contractor: facilitators & note taker)
Participants		
European Commission	SPC secretariat and large Commission delegation	Commission thematic expert, country desk, officer responsible for the programme
Member States	<ul style="list-style-type: none"> All EU Member States 	- Limited selection of peer countries (usually 7 or 8, up to 10 participating countries)
Composition of Member State delegation	<ul style="list-style-type: none"> SPC delegations 	- For every peer country two representatives (one government + one non-governmental expert)
Stakeholders	<ul style="list-style-type: none"> As a rule no participants external to the SPC, except (occasionally) some experts 	- A representative of 2 European stakeholder networks; host country stakeholders; a nongovernmental network expert
Number of participants	Usually more than 60 people	Usually between 30 and 40 people
Format	<ul style="list-style-type: none"> Plenary sessions to start and to end the in depth review Working Groups 	<ul style="list-style-type: none"> Plenary sessions In some cases: Working groups Site visits
Reporting	No or minimal formal reporting.	Extensive reporting on website: all papers that were used as input, short and synthesis report, minutes of the meeting.
Follow up	Feeds directly into the Joint Report through European Commission/SPC secretariat	Weaker / more indirect link to the Joint Report (main Social OMC messages)

³⁰ The PROGRESS peer review programme is not only open to EU Member States but also to accession countries and other European countries not part of the EU that adhere to the programme.

The best kept secret in Brussels: OMC 'Projects'

Since the start of the Social OMC, each year the Commission has launched a call for proposals inviting project proposals on social protection and social inclusion issues. Calls have been launched on awareness raising, mutual learning projects, social experimentation. Through these projects, the Commission is promoting some action, cooperation and exchanges 'in the field'. Projects are funded on a co-financing basis. They typically last between one and two years. For most calls there is an explicit condition that projects be proposed by a partnership covering several Member States, and partnerships of different types of actors are encouraged — NGO's, ministries, universities, and local government. Except for the national awareness-raising calls, in most cases the projects are actually benchmarking or peer review exercises where partners document policies in different countries and organise seminars in each of the participating Member States. The results of the projects are typically put on project-funded websites. They are not accessible from the European Commission's website, so dissemination is not as effective as it should be.

The quality of the projects is certainly uneven: some have been more successful than others. Although it would require thorough analysis to trace the impact of the projects themselves, it is clear that results from the more successful projects have fed back into the central process through the participating actors: NGO's, Member States etc. An example of a project that had a wider impact is briefly described in box 7.3.

Box 7.3 Example of OMC Project: micro-finance exchange network

Developing an exchange network for European actors of micro-finance (transnational exchange project 2002).³¹ Partners in this project were *Association pour le Droit à l'Initiative Economique* (France), Evers & Jung — Research and Consulting in Financial Services (Germany) and the New Economics Foundation (United Kingdom). The goal of this six month project was the promotion of the mutual learning and the exchange of know-how, as well as the dissemination of good practices between the microcredit operators, in order to improve the microcredit impact quality and effectiveness in terms of action against social exclusion in Europe (action in favour of self-employment of people in need). This project led to the launch of the European Microfinance Network in April 2003 (21 members). This network later developed into a core funded EU stakeholder network, which currently counts 93 member organisations in 21 European countries.

³¹ See http://www.european-microfinance.org/index2_en.php.

Drawing conclusions in Joint Reports on Social Protection and Social Inclusion and Commission Recommendations

Joint reports bring together the lessons learned as a result of the mutual learning/benchmarking activities during the year. A draft of the report is prepared by the Commission at the end of the year, discussed with Member States and EU stakeholder networks, and a final version approved by the Social Affairs Council at the start of the next year. A joint report consists of two parts: the actual joint report (where 'joint' means approved by both the European Commission and the Council) which is rather concise (typically 10–15 pages), and the more elaborate supporting document published as a Commission Staff Working Document (although the subject of discussions between the Commission and Member States, the final responsibility lies with the Commission). The joint report contains 'Key Messages' (merely a few pages) that are communicated by the Social Affairs Council to the Spring European Council, the annual meeting of EU heads of state and government that takes stock of the state of the EU.

The supporting document contains an analysis of the commonly agreed indicators (the social situation of the EU) or an in-depth analysis of the key issue in thematic years on the basis of the commonly agreed indicators (presenting graphs which rank Member States — see section 7.3, *Indicators: genuine benchmarking in spite of sensitivities*). Then there is the policy analysis part, which presents a horizontal analysis (across Member States) of the major social protection and inclusion issues reviewing challenges and Member States' policies. Finally, there are country fiches, which summarise the results of the vertical analysis per Member State. Although the EU treaty does not provide for individual country recommendations on social protection and social inclusion (in this respect there is a difference with the European Employment Strategy) the country fiches do identify country specific 'challenges'.

Even though the academic literature has classified these messages from the Commission as too general to have any real bite, box 7.4 below illustrates that in some cases suggestions to the Member States can be rather far-reaching, and really represent 'soft recommendations' which cover the three strands of the Social OMC.

Because the joint report is a rather unwieldy document that summarises lessons learned on many issues on an annual basis it is not very accessible for people looking for guidance on a specific issue. This may partly explain why in general it is only known within the OMC community. The Commission has proposed to summarise lessons learned on key issues (the results of the consensus framing exercise) in Commission recommendations (CEC 2008b, p. 5). Through such Commission recommendation, results of the mutual learning exercise can be

brought together and made visible. Since the recommendation is subsequently discussed by the other European institutions, this makes it possible to involve EU institutions not normally involved in the Social OMC, like the European Parliament. The follow up, monitoring of the impact of the recommendation can then be organised within the Social OMC. Thus, a Commission recommendation on the active inclusion of people excluded from the labour market was published in October 2008 (CEC 2008a). In December 2008 the EPSCO council endorsed the recommendation (Council of the EU 2008), and in May 2009 the European Parliament endorsed it (EP 2008).

Box 7.4 Examples of country specific challenges (subtle ‘recommendations’) through the Joint Reports

Governance

Austria: While the presented objectives are by themselves very important, they are in major parts not made more concrete by target setting and rolling out a financial perspective to underpin the process (CEC 2004).

Social inclusion

Spain was advised that ‘Co-ordination and co-operation between the different administrative levels will be required to define a minimum standard of measures in order to tackle the inclusion issue in a more homogenous way throughout the national territory’ (CEC 2002).

Pensions

The *Czech Republic* was asked to ‘To ensure that reforms (e.g. privatization of funds) are properly thought through on the basis of past experience and the experience of other countries’ (CEC 2009b).

Health care

Luxembourg was asked to limit ‘the overuse of antibiotics’ and improve the use of generic medicines (with regard to quality and financial sustainability) (CEC 2007; CEC 2009b).

7.4 Europe 2020 Strategy: is there still room for social benchmarking?

While the previous section focussed on social benchmarking throughout the past decade, this section briefly introduces the main features of the Europe 2020 Strategy and discusses the key changes this entails for benchmarking within the Social OMC.

A new strategy for Europe until 2020: objectives and governance

Overall objectives and general architecture of the new strategy

In the words of the European Commission (CEC 2010b), Europe 2020 is a Strategy ‘to turn the EU into a smart, sustainable and inclusive economy delivering high levels of employment, productivity and social cohesion’. This new Strategy is based on enhanced socio-economic policy coordination, and is organised into three priorities, which are expected to be mutually reinforcing:

- *smart growth* — ‘strengthening knowledge and innovation as drivers of our future growth’
- *sustainable growth* — ‘promoting a more resource efficient, greener and more competitive economy’
- *inclusive growth* — ‘fostering a high-employment economy delivering social and territorial cohesion’.

Europe 2020 has been organised around three (supposedly integrated) pillars:

- *Macroeconomic surveillance*, which aims at ensuring a stable macroeconomic environment conducive to growth and employment creation. This is the responsibility of the EU ‘Economic and Financial Affairs’ (ECOFIN) Council.
- *Thematic coordination*, focusing on structural reforms in the fields of innovation and R&D, resource-efficiency, business environment, employment, education and social inclusion. Thematic coordination combines EU priorities, EU headline targets (translated into national targets that underpin them) and EU flagship initiatives. It is conducted by the sectoral formations of the EU Council of Ministers. This includes, for social protection and inclusion matters, the EPSCO Council. In other words, this is where we find the Social OMC in the new architecture.
- *Fiscal surveillance under the Stability and Growth Pact*, which should contribute to strengthening fiscal consolidation and fostering sustainable public finances.

*Europe 2020: the EU and domestic level*³²

The June 2010, the European Council agreed to set ‘five EU headline *targets* which will constitute shared objectives guiding the action of Member States and the Union’ (European Council 2010b). These include an EU target that aims *to promote*

³² This section draws on Frazer *et al.* (2010) and Vanhercke (2011).

*social inclusion, in particular through the reduction of poverty, by aiming to lift at least 20 million people out of the risk of poverty and exclusion. The target will consist in reducing the number of people in the EU who are at risk of poverty and/or materially deprived and/or living in jobless households (120 million) by one sixth.*³³ With a view to meeting these EU-wide targets, Member States then have to set their *national* targets, taking account of their relative starting positions and national circumstances and according to their national decision-making procedures. They should also identify the main bottlenecks to growth and indicate, in their National Reform Programmes (NRP's), how they intend to tackle them. Member States submitted their second NRP's in April 2012.

To underpin these targets and 'catalyse progress under each priority theme', the Commission has also proposed seven *flagship initiatives* that should encompass a wide range of actions at national, EU and international levels: 'Innovation Union', 'Youth on the move', 'A digital agenda for Europe', 'Resource efficient Europe', 'An industrial policy for the globalisation era', 'An agenda for new skills and jobs' and 'European platform against poverty' (EPAP) (CEC 2010c).

Finally, ten *Integrated Guidelines* for implementing the Europe 2020 Strategy were adopted by the Council in October 2010 — six broad guidelines relating to the economic policies of the Member States and the EU, and four guidelines concerning the employment (and in fact also social) policies of the Member States. Their aim is to provide guidance to Member States on defining their NRPs and implementing reforms, in line with the Stability and Growth Pact. All of these instruments are brought together into a time frame that has been labelled the 'European Semester' (see Frazer, *et al.* 2010 for a more detailed discussion). The question then is whether this new framework has important ramifications for social benchmarking.

Benchmarking Social Europe under the new strategy: a blueprint for a reinvigorated Social OMC: opportunities and uncertainties

As indicated in the introduction, the Social Affairs Ministers of the 27 Member States decided on a blueprint that outlines the main features of the future Social OMC. They did so by endorsing (in the EU jargon) an Opinion that was prepared by the Social Protection Committee (Council of the EU 2011).

Importantly, the SPC's blueprint for the future Social OMC starts by reaffirming that the Social OMC will continue to work in a 'holistic' way in that it will:

³³ The other targets relate to employment, education, research and development (R&D) and climate/energy.

-
1. Cover the three strands of the Social OMC — inclusion, pensions *and* health care and long-term care.
 2. Encompass adequacy, financial sustainability *and* modernisation of social protection systems.
 3. Be an essential tool to support the Social Affairs Ministers in monitoring and assessing the *entire* social dimension of the Europe 2020 Strategy.

In other words: Social Affairs Ministers seem to see the OMC as a way to reclaim some of the territory they lost through a narrow focus of the Europe 2020 Strategy on ‘Growth and Jobs’. This seems to be confirmed by the affirmation that while ‘The Social OMC ... contributes to the Europe 2020 governance cycle (European Semester)’, it does so ‘while maintaining its specificity’ (read: its independence).

What are the main **features, questions and pitfalls** of the blueprint for the Social OMC in the next decade?

A general concern is whether in the longer run the **Social OMC will remain viable**. In view of important capacity constraints of both the dedicated Commission services and the Member States, there is considerable pressure to concentrate on one central process: Europe 2020.

It seems that a wide consensus exists around the **Common Objectives** on Social Protection and Social Inclusion: the existing set is reconfirmed. As indicated before, the main value of the objectives lies in their balanced approach to the social policy areas. They drive home the point that in order to be effective policies need to simultaneously reach different objectives: most importantly financial sustainability *and* social adequacy. The test will be whether such balanced policies will be a central characteristic of high level EU policy initiatives in the coming years. Arguably, the objectives have been somewhat strengthened through so-called ‘minor technical updating’, which includes ‘taking full account of the relevant social provisions of the Lisbon Treaty’. One of the new social provisions is the so called horizontal social clause (article 9 TFEU)³⁴. The reference to the Lisbon Treaty (and thus article 9) may provide additional legitimacy to the European Commission to intensify its efforts as regards the social component of its impact assessments (mainstreaming social protection and inclusion concerns in all relevant policy areas). So far, however, there are no signs that the European Commission has picked up on this.

³⁴ This article states that ‘In defining and implementing its policies and activities, the Union shall take into account requirements linked to the promotion of a high level of employment, the guarantee of adequate social protection, the fight against social exclusion, and a high level of education, training and protection of human health’.

Then, there is the ‘**national reporting**’, which so far happened through the National Strategic Reports (and National Action Plans on Social Inclusion). Here, the Council considers that the policy coordination process carried out under the Social OMC requires ‘regular strategic reporting’ covering policies and measures in the three strands of the Social OMC. The Member States are currently submitting their first National Social Reports, new style. The reports are supposed to be coherent with and complementary to the NRPs.

The move to annual reporting (until now NSR’s were presented every three years) gives rise to some hope that the NRPs will be underpinned by comprehensive national strategies on social protection and social inclusion, and that the key elements of these will be referenced in the NRPs, but much will depend on the actual implementation: it seems more likely that the wording could lead to minimalistic interpretations of Member States’ reporting obligations.

The annual **Joint Reports** (adopted by the Council and the European Commission) have been abolished. They have been replaced by annual SPC Reports on the Social Dimension of the Europe 2020 strategy that are no longer formally adopted by the European Commission as a ‘college’ (decisions are deliberated collectively and the Members have collective responsibility). This raises concerns regarding a possible downgrading of the social assessment and a further reduction of its status, visibility and potential impact. The SPC opinion on the Social OMC further specifies that ‘The thematic work of the SPC *could* also lead to joint reports of the Commission and of the Member States’ (our italics). It remains to be seen whether such thematic Joint reports will indeed be forthcoming. A Commission recommendation on child poverty has been announced for 2012.

In line with the ‘holistic’ approach to the continued OMC, the decision has been taken to continue working on **indicators** in all three strands, while ensuring consistency with the newly established Joint Assessment Framework.³⁵ Among the priorities for indicator development in the coming years recently identified by the SPC are: the monitoring of interactions between social, economic, employment, and environmental policies and addressing the situation of the most vulnerable groups, such as the Roma (SPC 2011, p. 8). Again, the actual implementation of this blueprint will have to clarify what is meant with the Council’s open-ended commitments such as: ‘Member States and the Commission should continue to draw *selectively, as need be*, on the full set of commonly agreed indicators’ (our emphasis) in the context of the new reporting mechanisms.

Peer reviews are reconfirmed by the Council and the SPC as essential feature of the Social OMC, but they are to be more closely linked to policy reforms and the

³⁵ See: <http://ec.europa.eu/social/main.jsp?langId=en&catId=89&newsId=972&furtheNews=yes>.

dissemination of outcomes is to be improved (here reference is made to ‘Progress peer reviews’). Yet again, a key question prevails as regards the future of this tool, notably whether additional innovative peer review formats that have already been discussed in the SPC — in order to answer some shortcomings in the current mutual learning exercise — will be taken up. Among these formats are Social OMC workshops that could be a broader intermediary forum between individual peer reviews and the Social Protection Committee and follow up seminars in the host countries that would allow discussing lessons learned with a broader range of domestic stakeholders, notably also people involved in policy implementation at the local level.

While the SPC considers it important to improve the **involvement of stakeholders** such as social partners, NGOs, regional and local authorities as well as other relevant EU institutions, no specific proposals have been envisaged so far.

In other words, the proof of the pudding as regards the future of the Social OMC (and benchmarking as an essential part thereof) will be in the eating, as can also be seen from box 7.5 which summarises the key features, questions and pitfalls of the ‘reinvigorated’ Social OMC.

Box 7.5 Blueprint of the Social OMC under the Europe 2020 Strategy: key features, questions and pitfalls

1. **Common Objectives** — existing Objectives remain valid, but with ‘technical update’ (coherence with Europe 2020 and especially social provision of Lisbon Treaty); *Question*: will the European Commission enhance Social Impact Assessments?
2. **National reporting** — succinct annual reporting on social protection and social inclusion. *Question*: will key elements of comprehensive SPSI strategies be referenced in the NRPs? *Risk*: minimalistic interpretations of Member States’ reporting obligations).
3. **Indicators** — continue work on indicators in all three strands. *Risk*: Member States ‘draw too selectively’ on commonly agreed indicators in reporting.
4. **Joint Reports** — only published on specific occasions (thematic reporting). Replaced by ‘Annual Report of the SPC’. *Risk*: downgrading of social assessment and further reduction of status, visibility and impact.
5. **Peer Reviews** — reconfirmed as essential learning tools but should be more closely linked to policy reforms. *Question*: will innovative seminar formats linked to peer reviews be implemented? How to improve visibility and accessibility?
6. **Stakeholder involvement** — reconfirmed as important objective. *Question*: readiness to work on ‘voluntary guidelines’ as regards good quality stakeholder participation? *Risk*: further deterioration of participation as consequence of lighter reporting.

7.5 Wrapping things up: benchmarking in the social OMC demystified?

This chapter started off by describing the political setting in which the OMC emerged and was further institutionalised: in some politically sensitive areas where the unanimity rule prevented EU legislative initiatives, decision makers agreed that ‘doing nothing’ at EU level was not an option. As a result, the OMC emerged as a form of ‘governance by objectives’ through which Member States wanted to learn from one another (and thereby improve policies), while affirming that European integration equally had a visible social face. Since 2000, OMC-type processes and approaches have been proposed and applied by European institutions and stakeholders as mechanisms for coordinating domestic policies in a range of issue areas for which the EU has no formal authority, but also for monitoring and supplementing EU legislative instruments.

In the absence of a ‘shadow of hierarchy’ (Héritier and Lehmkuhl 2008), this ‘soft’ coordination tool seems hardly *adequate* to have any ‘real’ effect, at least this is the view of the ‘sceptical’ literature which by and large dismissed the OMC because of its institutional weakness. And yet the more ‘optimistic’ academic literature finds that at least the Social OMC (which is one of the more developed coordination processes) has a significant *impact* on domestic and EU policies, namely through mechanisms such as leverage and policy learning (see, for example, Vanhercke 2010).

In an attempt to unravel these contradictory findings, we set off to describe the way benchmarking really works in the context of the Social OMC, as it was clear to us that students of the OMC *are often* rather vague about *what they are actually talking about*, when they discuss the success and (especially) failure of ‘benchmarking’ in this context. By doing so, we found that Social OMC benchmarking is quite different from how it is usually portrayed in the literature in at least five respects.

First, it seems that the OMC’s learning tool are **more dynamic** than is usually acknowledged: thus, we described how the Peer Review mechanism was refined over the years, from (early) peer reviews on comprehensive national social inclusion plans to (later) in depth reviews of specific subjects or policies. We also saw how the set of indicators became increasingly multidimensional while covering more issue areas, development of input and output statistics; and we pointed to the increased use of quantified objectives or targets.

Second, the benchmarking tools are **more diversified** than is usually depicted, as we found when describing ‘SPC-peer reviews’ versus ‘PROGRESS’ Peer Reviews,

while for the latter two modalities can be identified: good practice peer reviews and policy problem or policy reform peer reviews. The latter distinction is more than just a nuance: the policy reform peer reviews are by no means the ‘window dressing’ exercises that are sometimes caricatured in the literature: discussions are often quite critical towards the presented policy, also because ‘site visits’ allow for a certain reality check.

Third, our in-depth description of social benchmarking in the Social OMC tool suggests that its tools have **more bite** than is usually assumed. OMC might look ‘soft’ but in some cases it feels quite hard to those who are touched by it. The same is true for the considerable pressure on Member States to start working with targets (even if this does not sit easily with their national culture). We presented a number of examples (*e.g.* child poverty) where Member States are ranked, not only as regards their outcomes, but also regarding the main causal factors of these outcomes. While country-specific messages are often deemed ‘too subtle’ even to be assessed by scholars, we provided some illustrations of ‘challenges’ for Member States identified through the JR SPSI, which in some cases are rather far-reaching, and really represent ‘soft recommendations’ which cover the three strands of the Social OMC. In other words, it is clear that ‘genuine’ benchmarking is going in the Social OMC, in spite of sensitivities, and that the ‘hard politics of soft law’ is not fiction.

Fourth, it seems that the OMC benchmarking is **more open** (involves a wider range of actors) than oftentimes assumed. In other words, the image of the Social OMC (especially but not only peer reviews) as involving a rather closed circle of non-accountable bureaucrats (Kröger 2009, p. 13) is sometimes exaggerated. Thus, the existence of supporting tools such as the nongovernmental expert and EU (civil society) stakeholder networks, are rarely acknowledged. And yet, through the exchanges with other networks members and their participation in drawing up EU comparative analyses (for example, benchmarking exercises, and development of scorecards), the experts acquire cross-country knowledge and become part of what could be called the ‘Social OMC community’. They can have an important impact on the quality of the benchmarking exercise. The same is true for the actors participating in the described (yet well-hidden) OMC ‘projects’: over the years the projects have equally supported the development of the Social OMC community, through the introduction of new partners to the process and by supporting the development of the EU civil society stakeholder networks. These stakeholders contribute to ‘capillary effects’ of the Social OMC in that benchmarking as a governance tool is increasingly being promoted outside the formal OMC inner circle (mainly governments and Commission officials).

It will be clear from this chapter that these capillary effects are by no means automatic processes: they are purposefully constructed by key actors in the process, for example, through funding. The crux of the matter is this: it is not the ‘hardness’ or the ‘softness’ of the OMC that matters, but its capacity to stimulate policy learning and benchmarking (through a rather refined set of tools) and especially creative appropriation and action by European, national and sub-national actors. Indeed, OMC benchmarking can only have an impact if it is being ‘picked up’ by actors at the domestic level, who use it as leverage to (selectively) amplify national reform strategies.

Thus, social policymakers are now facing the key challenge of re-establishing a role for a wide range of actors within the heart of the Social OMC, including by implementing innovative peer review formats and dissemination practices, developing guidelines for the quality involvement of stakeholders. And exploring the possibility of introducing procedural requirements into soft law mechanisms, such as rights to transparency and participation in the OMC. If they succeed, these policymakers (and the larger OMC community) may very well reach the goal they wanted to achieve by reinvigorating the Social OMC in June 2011: to counterbalance the Europe 2020 Strategy’s excessive focus on fiscal and economic considerations in the first so called ‘European Semester’ and the narrowing down of social policy to policy against poverty and social exclusion in the set up of the new strategy (for example in the European Commission’s first Annual Growth Survey). One way to provide such counterweight would be to continue producing sound analysis and political messages about how broad ‘social protection (social insurance) and social inclusion’ strategies are to be considered as a productive factor.

By pursuing this avenue, the EU would live up to its own commitment to deliver ‘inclusive growth’ and perhaps even provide Europe’s people with hope for an ever stronger Social Europe.

References

Börzel, T. 2003, *What can Federalism teach us about the EU? The German experience*, The Federal Trust for Education & Research, online paper 17.

Caritas Europa 2007, *An Interim Assessment of Europe’s National Social Inclusion Strategies*, p. 3, Brussels, <http://www.caritas-europa.org/module/FileLib/CONCEPTENLowRes.pdf>.

CEC 2007, *Joint Report on Social Protection and Social Inclusion 2007*.

——— 2004, *Joint Report on Social Inclusion 2004*.

-
- 2002, *Joint Report on Social Inclusion 2002*.
- 2008a, Commission Recommendation 2008/867/EC of 3 October 2008 on the active inclusion of people excluded from the labour market, *Official Journal of the European Union* L 307/11 of 18 November 2008, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:307:0011:0014:EN:PDF>.
- 2008b, Commission Communication ‘A Renewed Commitment To Social Europe: reinforcing the open method of coordination for social protection and social inclusion’, COM (2008) 418, Brussels, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0418:FIN:EN:DOC>.
- 2009a, *Portfolio of indicators for the monitoring of the European strategy for social protection and social inclusion*, <http://ec.europa.eu/social/BlobServlet?docId=3882&langId=en>.
- 2009b, *Joint Report on Social protection and Social Inclusion 2009*, 524 pages.
- 2010a, Annual Growth Survey: Advancing the EU's Comprehensive Response to the Crisis, Communication from the Commission, COM (2011) 11, Brussels, http://ec.europa.eu/europe2020/pdf/en_final.pdf.
- 2010b, Communication from the Commission ‘Europe 2020: A strategy for smart, sustainable and inclusive growth’, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF>.
- 2010c, Commission Communication ‘The European Platform against Poverty and Social Exclusion: A European framework for social and territorial cohesion’, COM (2010) 758, Brussels, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52010DC0758:EN:NOT>.
- 2010d, *Joint Report on Social protection and Social Inclusion 2010*, 304 pages, <http://ec.europa.eu/social/BlobServlet?docId=4667&langId=en>.
- Council of the EU 2006, Endorsement of the Joint Social Protection Committee/Economic Policy Committee Opinion on the Communication from the Commission ‘Working together, working better: A new framework for the open coordination of social protection and inclusion policies in the European Union’; Council of the European Union document 6801/06 pp. 8–10 corrigendum 5/07/2006.
- 2008, ‘Council Conclusions on common active inclusion principles to combat poverty more effectively, 2916th Employment, Social Policy, Health and

-
- Consumer Affairs Council meeting', Brussels, http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/lsa/104818.pdf.
- 2011, 'Endorsement of the Opinion of the Social Protection Committee on reinvigorating the Social OMC in the context of the Europe 2020 Strategy', doc. 10405/11, SOC 418, Brussels, <http://register.consilium.europa.eu/pdf/en/11/st10/st10405.en11.pdf>.
- de Búrca, G. and Scott, J. 2006, 'Introduction: New Governance, Law and Constitutionalism' in G. de Búrca and J. Scott (eds), *Law and New Governance in the EU and US*, Hart, Oxford.
- EP 2008, Active inclusion of people excluded from the labour market. European Parliament resolution of 6 May 2009 on the active inclusion of people excluded from the labour market (2008/2335(INI)), <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=//EP//NONSGML+TA+P6-TA-2009-0371+0+DOC+PDF+V0//EN>.
- and Council of the EU 2006b, Decision No 1672/2006 of the European Parliament and of the Council of 24 October 2006 establishing a Community Programme for Employment and Social Solidarity — PROGRESS, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:315:0001:0008:EN:PDF>.
- and — 1999, Regulation (EC) No 1784/1999 of the European Parliament and of the Council of 12 July 1999 on the European Social Fund, OJ L213/5 of 13 August 1999, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31999R1784:EN:HTML>.
- Espeland, W.N. and Stevens, M.L. 1998, 'Commensuration as a Social Process', *Annual Review of Sociology*, vol. 24, pp. 313–343.
- European Council 2000, *Presidency Conclusions – Lisbon European Council 23 and 24 March 2000*, http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100r1.en0.htm.
- 2010b, *Conclusions – 17 June 2010*, EUCO 13/10, http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/115346.pdf.
- Frazer, H., Marlier, E., Natali, D., Van Dam, R. and Vanhercke, B. 2010, 'Europe 2020: Towards a More Social EU?', in Marlier, E. and Natali, D. (eds.), *Europe 2020: towards a more social EU?*, P.I.E. Lang, Brussels.

-
- Greer, S. and Vanhercke, B. 2010, 'Governing Health Care through EU Soft Law', in E. Mossialos, T. Hervey, R. Baeten (eds), *Health System Governance in Europe: the role of EU law and policy*, Cambridge University Press, Cambridge.
- Hemerijck, A. and Visser, J. 2001, 'Learning and Mimicking: how European welfare states reform', typescript.
- Héritier, A. and Lehmkuhl, D. 2008, 'The Shadow of Hierarchy and New Modes of Governance', *Journal of Public Policy*, vol.28, no. 1, pp. 1–17.
- Kröger, S. 2009, 'The Open Method of Coordination: underconceptualisation, overdetermination, de-politicisation and beyond', in S. Kröger (ed.) *What Have We Learnt: advances, pitfalls and remaining questions in OMC research*, European Integration online Papers (EIoP), Special Issue 1, Vol. 13, Art. 5, <http://eiop.or.at/eiop/texte/2009-005a.htm>.
- Laursen, F. 2011, *The EU and Federalism: politics and policies compared*. Farnham, Ashgate.
- Leibfried, S. and Pierson, P. 1995, 'European Social Policy. Between fragmentation and integration', Brookings Institute, Washington DC.
- Majone, G. 2006, 'Federation, Confederation, and Mixed Government: a EU–US comparison', in Anand Menon and Martin Shain, (eds), *Comparative Federalism: the European Union and the United States in comparative perspective*, Oxford University Press, New York.
- Meyer, H., Barber, S. and Luenen, C. 2011, The Open Method of Coordination. A Governance Mechanism for the G20? Bertelsmann Stiftung, p. 21, http://www.bertelsmann-stiftung.de/bst/en/media/xcms_bst_dms_34910_34911_2.pdf.
- Moravcsik, A. 2001, 'Federalism in the European Union: Rhetoric and Reality', in *The Federal Vision*, pp. 161-189, Oxford Scholarship Online Monographs, Oxford.
- OSI 2010, No Data—No Progress: country findings, data collection in countries participating in the Decade of Roma Inclusion 2005–2015, Open Society Institute, <http://www.unhcr.org/refworld/docid/4cbc430e2.html>.
- SPC 2008, Child Poverty and Well-being in the EU – Current Status and Way Forward, OPOCE, Luxembourg. <http://ec.europa.eu/social/BlobServlet?docId=2049&langId=en>.
- 2011, Background Paper to the SPC from the ad hoc group on reinvigorating the Social OMC in the context of the Europe 2020 Strategy, Doc. SPC /2011.09/1, p. 31.

-
- and CEC 2010, 2010 Update of the Joint Assessment by the SPC and the European Commission of the social impact of the economic crisis and of policy responses, Doc 16905/10, SOC 793, Brussels, <http://register.consilium.europa.eu/pdf/en/10/st16/st16905.en10.pdf>.
- Streeck, W. 1996, 'Neo-Voluntarism: a new European social policy regime?', in Marks, G. *et al.* (eds), *Governance in the European Union*, SAGE, London.
- TFEU 2010, Consolidated version of the Treaty on European Union (TEU) and the Treaty on the Functioning of the European Union (TFEU), *Official Journal of the European Union*, C 83, vol. 53, no. 30, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0047:0200:en:PDF>.
- Vandenbroucke, F. 2001, 'New Policy Perspectives for European Cooperation in Social Policy', Speech at the European Conference 'Social and Labour Market Policies: Investing in Quality', Brussels, <http://oud.frankvandenbroucke.be/html/soc/T-010222bis.htm>.
- Vanhercke, B. 2010, 'Delivering the Goods for Europe 2020? The Social OMC's adequacy and impact re-assessed', in Marlier, E. and Natali, D. (2010), *Europe 2020: Towards a more Social EU?*, Peter Lang, Brussels.
- 2011, 'Is "The Social Dimension of Europe 2020" an Oxymoron?', in Degryse, C. and Natali, D. (eds), *Social Developments in the European Union 2010*, ETUI, OSE, pp. 141–74, Brussels.
- Verschraegen, G., Vanhercke, B. and Verpoorten, R. 2011, 'The European Social Fund and Domestic Activation Policies: europeanization mechanisms', *Journal of European Social Policy*, vol. 21 no.1, pp. 1–18.
- Wallace, H. and Wallace, W. 2001, *Policy Making in the European Union*, Oxford University Press, Oxford.
- Zeitlin, J. 2005, 'The Open Method of Co-ordination in Question', in Zeitlin, J. & Pochet, P. (eds), *The Open Method of Co-ordination in Action: the European Employment and Social Inclusion Strategies*, P.I.E.-Peter Lang, Brussels.

PART II

AUSTRALIAN CONTRIBUTIONS

8 Australia's federal context*

Gary Banks¹

Productivity Commission

Steering Committee for the Review of Government Service Provision

Alan Fenna²

Curtin University

Lawrence McDonald³

Productivity Commission

Secretariat for the Review of Government Service Provision

8.1 Australian federalism

Australian federalism has evolved since its beginnings over a century ago into a system where the Commonwealth government is engaged in a wide range of policy areas that were once the sole responsibility of the States. This system of 'cooperative federalism' has prompted efforts to establish new and more efficient and effective modes of intergovernmental coordination. At the centre of those efforts has been COAG, the Council of Australian Governments. At issue has been the operation of the extensive system of 'tied' grants through which the Commonwealth shapes policy in areas of State jurisdiction. While the States retain primary responsibility for most service delivery, they increasingly do so within the context of an overarching national framework.

* The views expressed in this chapter are those of the authors, and do not necessarily represent the views of the Steering Committee for the Review of Government Service Provision.

¹ Gary Banks is the Chairman of the Productivity Commission and Chairman of the Steering Committee for the Review of Government Service Provision.

² Alan Fenna is Professor of Politics at Curtin University.

³ Lawrence McDonald is an Assistant Commissioner at the Productivity Commission and Head of the Secretariat for the Steering Committee for the Review of Government Service Provision.

The constitutional division of powers

The division of powers in Australia's Constitution was deliberately decentralised. The Commonwealth was assigned limited and specific powers, whereas State powers are general (box 1). The intention and expectation at federation was that the States would retain exclusive responsibility for most domestic governance tasks and that little coordination would be required between the two levels of government (Fenna 2007b). In particular, most of the service delivery responsibilities of government in areas such as education, health and infrastructure as well as regulatory responsibilities in areas such as land use and the environment, were left exclusively to the States. The two levels were assigned concurrent jurisdiction in respect of all forms of taxation except customs and excise, which were prohibited to the States.

Box 8.1 Australian and State government division of powers

The Australian Constitution assigns the Australian Government:

- a small number of exclusive powers — mainly in respect of customs and excise duties, the coining of money and holding of referenda for constitutional change ; and
- a large number of areas where it can exercise powers concurrently with the States. To the extent that State laws are inconsistent with those of the Commonwealth Government in these areas, the laws of the Commonwealth prevail (s.109).

State governments have responsibility for all other matters. However, even where the Constitution does not give the Commonwealth explicit power, it may be able to draw on more general powers, such as the 'corporations' power and the 'external affairs' power. Further, the Australian Government can influence State policies and programs by granting financial assistance on terms and conditions that it specifies (s.96).

Sources: PC 2006a, 2006b; Fenna 2007b.

Centralisation

In practice, the distribution of powers has become significantly more centralised over time, even though the Constitution itself remains largely unchanged (Fenna 2007a, 2012). Since federation, there has been an expansion in the role of government in general, and of the Commonwealth's role in particular. For the most part, increasing centralisation of power in Australia has occurred through expansive interpretation of the Commonwealth's enumerated powers by the High Court (s.51). These decisions have reduced State taxing powers; allowed the Commonwealth to deploy its 'spending power' to direct the States; and given very broad scope to the

Commonwealth's enumerated powers areas as that over 'external affairs' and the 'corporations power'.

8.2 Federal financial relations

Pivotal to the relationship between the Commonwealth and the States in Australian federalism is the very different financial position of the two levels of government, the dependence of the States on grants from the Commonwealth, and the capacity that financial superiority gives the Commonwealth (Fenna 2008).

Vertical fiscal imbalance

Because the customs tariff that was assigned exclusively to the Commonwealth was such an important source of revenue at the time, Australia federalism began with a significant degree of 'vertical fiscal imbalance' (VFI). The Commonwealth's revenues exceeded its minor spending needs, whereas the extensive service delivery responsibilities of the States exceeded their revenues. Since then the disparity has increased. Over time the High Court interpreted the tariff and excise prohibition as encompassing any kind of State sales tax; and in 1942, the Commonwealth took the personal and corporate income tax from the States. The Commonwealth now controls approximately 82 per cent of all tax revenue raised in Australia. In 2007-08 (the most recent year for which consistent data are available), around 46 per cent of total State revenue was provided by the Commonwealth.

An extensive system of intergovernmental transfers has developed to redress the imbalance. The Commonwealth traditionally allocated funds to the States either as general purpose grants or specific purpose payments (SPPs) — that is to say, either with no particular strings attached or as 'tied grants' carrying any of a range of particular conditions.

Table 8.1 Estimated State revenue, by source, 2009–10

	<i>\$ billion</i>	<i>% of total</i>
Own source revenue		
Tax	55	28
Other ^a	43	22
<i>Subtotal</i>	98	50
Commonwealth transfers		
GST revenue (untied)	45	23
National Agreement SPPS (loosely tied) ^b	24	15
National Partnership Payments ^c	28	13
<i>Subtotal</i>	98	50
Total	196	100.0

^a Includes sales of goods and services, regulatory fees and fines. ^b Specific Purpose Payments related to National Agreements must be spent in a nominated area, but do not have prescriptive conditions. ^c National Partnership Payments can include prescriptive conditions on how money is spent. **GST** Goods and Services Tax.

Source: ABS (2011), Government Finance Statistics, Australia, Cat. no. 5512.0.

Intergovernmental grants from the central government are characteristic of virtually all federations, even where sub-national governments have access to a range of revenue sources. Having the collection of some revenue sources centralised and others decentralised can reduce the overall cost of raising tax revenue and avoid competitive erosion of efficient tax bases (Pincus 2008). However, the degree of vertical fiscal imbalance in Australia has gone well beyond what such logic would suggest and well beyond the practice in comparable federations (Fenna 2008; Warren 2006).

VFI has become an issue, with many commentators noting the weakening of desirable links between taxation and expenditure decisions; the increased scope for the Commonwealth to become involved in areas of State responsibility (by attaching conditions to the use of transferred funds); and the heightening of political tensions around the allocation of revenue amongst the States (Garnaut and FitzGerald 2002). That said, replacement of the previous Commonwealth/State arrangements for untied grants with the proceeds of the GST ‘growth tax’ in 2000, and recent reforms to federal financial relations (discussed below) have shifted much of the focus of debate from the fiscal imbalance itself, to the mechanism for allocating Commonwealth transfers.

General purpose grants

Commonwealth grants to the States embody a significant degree of horizontal fiscal equalisation (HFE), a mechanism designed to ‘equalise’ the ability of State governments to deliver services. In 1933, following threats by Western Australia to

secede over claims of unfair financial treatment, the Commonwealth Grants Commission was established to advise on these special grants to the ‘weaker’ States. After the Commonwealth took over income tax from the States during the Second World War, the Grants Commission advised on the allocation to all States of shares of income tax. In 2000, the GST (Goods and Services Tax) was introduced, with all its net proceeds hypothecated to the States and the Grants Commission has been tasked with advising on the distribution of that revenue.

The Grants Commission describes the logic of HFE as follows:

State governments should receive funding from the Commonwealth such that, if each made the same effort to raise revenue from its own sources and operated at the same level of efficiency, each would have the capacity to provide services at the same standards. (CGC 2004)

This system has come under criticism at various times (Garnaut and FitzGerald 2002), most recently with the imbalances created by the mining boom (Fenna 2011; GST Distribution Review 2011; Porter 2011).

Tied grants

For some decades now, a large proportion of the transfers from the Commonwealth to the States has come in the form of individual grants directed toward specific purposes and often with various ‘input’ conditions attached. Prior to the reforms of 2008–09, the number of these Specific Purpose Payments (SPPs) had grown rapidly, reaching around 100. They had also become increasingly prescriptive, most notoriously to the point where one SPP implemented by the Howard government required States to ensure that every school was flying the Australian flag. Since at least the mid-1990s, various inquiries and commissioned reports had recommended that the Commonwealth relinquish its ‘micro-managing’ role by replacing the myriad SPPs with a small number of block grants while at the same time asking the States to be more openly accountable for what they managed to achieve with those funds.

8.3 Cooperation and collaboration in Australia’s federation

As noted, a distinctive feature of Australia’s federation is that many functions are now shared, rather than being exclusive to one level of government. What was by design a ‘coordinate’ system with each level of government operating

independently in its own sphere has become in practice a system of ‘concurrent’ jurisdiction.

Competition among the States has seen the introduction of a range of policy innovations in fiscal affairs and service provision that have spread across jurisdictions. However, in some areas, competition has been more destructive than constructive (for example, ‘bidding wars’ for investment and erosion of efficient tax bases) and in many areas has led to ongoing diversity that has detracted from good national policy outcomes (regulations inhibiting mobility or scale, inter-jurisdictional externalities, excessive transactions costs). As a consequence, various mechanisms have developed to promote more cooperative and coordinated action in areas where reform on a national scale was generally seen as being important and beneficial (Painter 1998).

The architecture of intergovernmental relations in Australia

Even before federation, the leaders of the Australian colonies met regularly to discuss issues of mutual interest or concern. From 1990 to 1992, Australian governments met in a series of Special Premiers’ Conferences to discuss the then-Prime Minister’s plan for a ‘Closer Partnership with the States’. At the last Heads of Government meeting in 11 May 1992, first ministers agreed to give their meetings more formal status as the Council of Australian Governments (COAG).

COAG was designed to operate as the main intergovernmental forum to ‘initiate, develop and monitor the implementation of policy reforms that are of national significance and which require cooperative action by Australian governments’. The operations of COAG are complemented by a range of Ministerial Councils, restructured most recently as ‘Standing Councils’ or ‘Select Councils’ (COAG 2011), which facilitate consultation and cooperation in specific policy areas. Ministerial Councils are mandated to develop policy reforms for consideration by COAG, and oversee the implementation of agreed policy reforms.

Recent reform experience

The first wave of economic reform in Australia occurred within areas of Commonwealth responsibility in the 1980s, and involved the floating of the currency and reduction of barriers to foreign goods and capital. This in turn exposed performance problems in other parts of the economy, including inefficiencies in infrastructure industries dominated by public monopolies, anti-competitive regulation of many product and service markets, and rigidities in the labour market (Banks 2005).

From the late 1980s, some governments accordingly started to tackle these problems in a second wave of reform. However, an Independent Committee of Inquiry into Competition Policy in Australia (Hilmer *et al.* 1993) demonstrated that effective implementation of many of the reforms required a more coordinated approach, and in April 1995, COAG committed to the National Competition Policy (NCC 2010).

The National Competition Policy included a range of reforms to expose government business enterprises to competitive pressures; to provide third-party access to essential infrastructure services and guard against the possibility of overcharging by monopoly service providers; and a process for reviewing a wide range of legislation that restricted competition (PC 2005).

The National Competition Policy was a landmark achievement in nationally coordinated economic reform. COAG (2005) stated at the conclusion of its June 2005 meeting:

A collaborative national approach was the cornerstone of successful implementation of the NCP reform agenda. It drew together the reform priorities of the Commonwealth, States and Territories to improve Australia's overall competitiveness and raise living standards

Key success factors included the formal commitment by all governments to specified reforms, and provision for 'competition payments' by the Australian Government to the States and Territories where they achieved satisfactory reform progress. Although the payments were relatively small (particularly in the context of the economic and social benefits that the States gain from undertaking reform), many commentators regarded them as essential to the reform process (PC 2005, 2006).

A third wave of reform, the National Reform Agenda, was agreed to in broad terms by COAG in February 2006. It encompasses three streams:

- competition reform, continuing the successful reforms of the 1990s
- regulation reform, to reduce the red tape burden on business
- human capital reform to improve health, learning and work outcomes, and therefore raise labour force participation and productivity in the face of an ageing demographic structure.

Like the National Competition Policy, the National Reform Agenda was based on the premise that cooperation between different tiers of government would lead to better outcomes for Australians. The States argued strongly for payments similar to those associated with the National Competition Policy, on the basis that, as much of

the fiscal benefit of the reforms would accrue to the Commonwealth, the Commonwealth should share these benefits with the reforming jurisdictions (Department of Treasury and Finance 2006).

Ultimately, the 2008 reforms to federal financial relations re-introduced a version of incentive payments through National Partnership Payments to support the delivery of specified outputs or projects, to facilitate reforms or to reward those jurisdictions that deliver on nationally significant reforms (Commonwealth of Australia and States and Territories 2009).

Modelling by the Productivity Commission of the National Reform Agenda indicates that the gains from this ‘third wave’ of reform could potentially be greater than from the first and second waves, depending on the nature of the specific reforms and their budgetary costs (PC 2006c). However, many of the reforms involve significant complexities and uncertainties. This has ‘upped the ante’ on having good analysis based on good evidence to help avoid making mistakes on a national scale which previously would have been confined to particular jurisdictions (Banks 2008; 2009). One mechanism for generating such evidence is the COAG-commissioned Report on Government Services (see Banks and McDonald, this volume).

Federal fiscal reform

Central to any real reform of intergovernmental relations in Australia, though, was reform to SPPs. States generally resented SPPs as ‘mechanisms for the Commonwealth to pursue its own policy objectives in areas of State responsibility’ (Ward 2009) and criticism was widespread (for example, Warren 2006). Many were narrowly focussed, prescriptive and inflexible. They inhibited the innovation and efficiency that can come from decision making attuned to local circumstances. For example, SPPs that required matching State funding led to States investing significantly more resources than they considered appropriate on some activities.

In November 2008, COAG endorsed a new Intergovernmental Agreement on Federal Financial Relations as part of a sweeping reform of Australian federalism launched by the incoming Labor government (Fenna and Anderson 2012). This aimed to replace existing SPPs with a small number of much less prescriptive transfers (COAG 2009). The agreement ‘rolled up’ multiple SPPs into five broad SPPs covering schools, vocational education and training, disability services, healthcare and affordable housing (Treasury 2009; 2010). Each SPP was associated with a new National Agreement that contained objectives, outcomes, outputs and performance indicators for each sector, and clarified the respective roles and

responsibilities of the Commonwealth and the States in the delivery of services. (A sixth National Agreement, the National Indigenous Reform Agreement, is not associated with an SPP.) Importantly, the National Agreements do not prescribe how States are to use the money in the related SPPs (beyond requiring the money to be spent in the relevant service sector). Rather, governments' performance under the National Agreements is monitored and assessed by the COAG Reform Council — an independent body that reports to COAG.

As part of these reforms, COAG also agreed to a 'new' form of payment — National Partnerships — to fund specific projects and to facilitate and reward States that deliver on agreed reforms. As at mid-2010, there were around 142 such National Partnerships, and they accounted for over 50 per cent of total SPP payments in 2009-10. This 'proliferation', and their sometimes opportunistic or interventionist nature, has led to suggestions that the undesirable features of the old system are re-emerging (for example, O'Meara and Faithful 2012).

The Intergovernmental Agreement on Federal Financial Relations established the COAG Reform Council as the key accountability body within the COAG architecture. An independent review body, the Council reports directly to COAG on reforms of national significance that require cooperative action by Australian governments (see O'Loughlin, this volume). It is this move to a new focus on outcomes performance that has introduced benchmarking to Australian federalism in a systematic way.

References

- Banks, G. 2005, 'Structural Reform Australian-Style: lessons for others?', Productivity Commission, Melbourne.
- 2008, 'Riding the Third Wave: some challenges in national reform', Productivity Commission, Melbourne.
- 2009, 'Evidence-based policy making: what is it? How do we get it?', Productivity Commission, Melbourne.
- COAG 2005, 'Communiqué', Council of Australian Governments, Canberra.
- 2009 'Communiqué', Council of Australian Governments, Canberra.
- 2011, 'Communiqué', Council of Australian Governments, Canberra.
- Commonwealth of Australia and States and Territories 2009, *Intergovernmental Agreement on Federal Financial Relations*, Council of Australian Governments, Canberra.

-
- Department of Treasury and Finance 2006, 'National Reform Agenda: the case for sharing the gains', Government of Victoria, Melbourne.
- Fenna, A. 2007a, 'The Malaise of Federalism: comparative reflections on Commonwealth–State Relations', *Australian Journal of Public Administration*, vol. 66, no. 3, pp. 298–306.
- 2007b, 'The Division of Powers in Australian Federalism: subsidiarity and the single market', *Public Policy*, vol. 2, no. 3, pp. 175–94.
- 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 Paul Kildea, Andrew Lynch, and George Williams (eds.), *Tomorrow's Federation: reforming Australian government*, Federation Press, Leichhardt, NSW, pp. 26–42.
- and Anderson, G. 2012, 'The Rudd Reforms and the Future of Australian Federalism', in Gabrielle Appleby, Nicholas Aroney, and Thomas John (eds), *The Future of Australian Federalism: comparative and interdisciplinary perspectives*, Cambridge University Press, Cambridge, pp. 393–413.
- Garnaut, R. and FitzGerald, V. 2002, *Review of Commonwealth–State Funding: final report*, Committee for the Review of Commonwealth–State Funding, Melbourne.
- Hilmer, F., Rayner, G., Mark R. and Taperell, G. Q. 1993, *National Competition Policy*, Committee of Review of the Application of the *Trade Practices Act 1974*, Canberra.
- NCC 2010, 'National Competition Policy', National Competition Council, Melbourne.
- O'Meara, P. and Faithful, A. 2012, 'Increasing Accountability at the Heart of the Federation', in Paul Kildea, Andrew Lynch, and George Williams (eds), *Tomorrow's Federation: reforming Australian government*, Federation Press, Leichhardt NSW, pp. 92–112.
- Painter, M. 1998, *Collaborative Federalism: economic reform in Australia in the 1990s*, Cambridge University Press, Melbourne.
- PC 2005, 'Review of National Competition Policy Reforms', *Inquiry Reports* Productivity Commission, Melbourne.
- (ed.) 2006, *Productive Reform in a Federal System: roundtable proceedings*, Productivity Commission, Canberra.
- Pincus, J. 2008, 'Six Myths of Federal–State Financial Relations', Committee for Economic Development of Australia, Melbourne.

Treasury 2009, *Budget Paper No. 3: Australia's federal relations 2009–10*, Commonwealth of Australia, Canberra.

— 2010, *Budget Paper No. 3: Australia's federal relations 2010–11*, Commonwealth of Australia, Canberra.

Ward, G. 2009, 'New Generation Specific Purpose Payments, Sweeping reform across Government', Queensland Treasury Brisbane.

Warren, N. 2006, *Benchmarking Australia's Intergovernmental Fiscal Arrangements*, New South Wales Treasury, Sydney.

9 Benchmarking and Australia's Report on Government Services*

Gary Banks¹

Productivity Commission

Steering Committee for the Review of Government Service Provision

Lawrence McDonald²

Productivity Commission

Secretariat for the Review of Government Service Provision

9.1 The Review of Government Service Provision

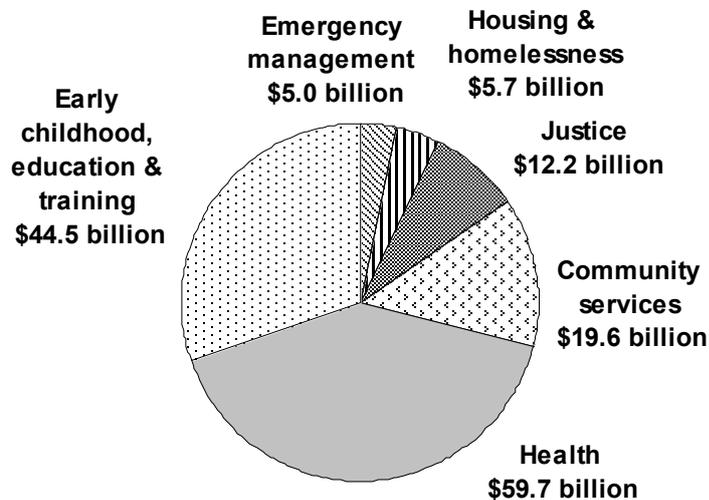
Every year Australia's governments cooperate in producing the Report on Government Services (RoGS), a comprehensive exercise in performance reporting across a wide range of services delivered by Australia's State and Territory governments. The range of services has grown since the first Report was published in 1995 and activities included in the 2011 Report amounted to almost \$150 billion, over two-thirds of total government recurrent expenditure, and were equivalent to about 12 per cent of Australia's gross domestic product (figure 9.1). It is a collaborative and consensual exercise in which the Commonwealth government plays a facilitative role rather than a directive or coercive one (see Fenna, this volume).

* A presentation to 'Benchmarking in Federal Systems: Australian and international experiences', a joint roundtable of the Forum of Federations and the Productivity Commission, Melbourne, Australia, 19 and 20 October 2010. *The views expressed in this paper are those of the authors, and do not necessarily represent the views of the Steering Committee for the Review of Government Service Provision.*

¹ Gary Banks is the Chairman of the Productivity Commission and Chairman of the Steering Committee for the Review of Government Service Provision.

² Lawrence McDonald is an Assistant Commissioner at the Productivity Commission and Head of the Secretariat for the Steering Committee for the Review of Government Service Provision.

Figure 9.1 Estimated government recurrent expenditure on services covered by the 2011 Report



Source: SCRGSP (2011), p. 1.8.

The Review of Government Service Provision (the Review) was established in 1993 by Heads of Government (now the Council of Australian Governments, or COAG) to provide comparative information on the efficiency, effectiveness and equity of government services across jurisdictions in Australia (SCRCSSP 1995). The Steering Committee's *Report on Government Services* (RoGS) commenced during what is now regarded as a transforming era of economic reform in Australia (Banks 2002).

During the 1980s and the 1990s, Australia underwent wide-ranging economic reform, including changes to monetary and fiscal policies, capital markets, industry assistance, taxation, regulation, labour markets and industrial relations, and innovation and training. These changes produced greater economic flexibility, improved efficiency and a more outward looking, opportunity-focussed business culture. They also yielded significant productivity dividends: through the 1990s productivity cycle, Australia's multi-factor productivity growth surged to an all-time high, averaging 2.1 per cent a year, three times our long-term average rate of 0.7 per cent (PC 2011).

Recognising the gains to the community from these extensive reforms within the 'private', or market, economy, governments realised that there were also large potential gains from improving the productivity of the public sector. But reform was challenging in areas for which there was no competitive market, and where criteria such as access and equity are particularly important. Australian governments recognised that the federation provided the opportunity to pursue reform by

comparing performance and learning from what other jurisdictions were doing and how they were doing it.

At their best, federal systems constitute a ‘natural laboratory’, in which different policy or service delivery approaches can be observed in action, providing the opportunity for learning about what works and what does not (see Fenna, this volume). Also, where one jurisdiction develops a successful new approach, other jurisdictions can adopt that approach at less cost than starting from scratch. However, taking advantage of diversity within a federal system requires an effective means of learning about and spreading successes — and, just as importantly, identifying and terminating failures (Banks 2005).

In 1991, Heads of Government accordingly requested the Industry Commission (predecessor of the Productivity Commission) to assist a Steering Committee of senior officials to set up a national system of performance monitoring for Government Trading Enterprises (GTEs) in the electricity, gas, water, transport and communication sectors (SCNPMGTE 1992). The resulting series of reports, known as the ‘red books’, stimulated substantial GTE reform, with significant economic pay-offs. The sweeping nature of these reforms, including the privatisation of many GTEs, ultimately led the Steering Committee to recommend its own disbandment in 1997 — although some further monitoring of the performance of GTEs has been conducted by the Productivity Commission as part of its general research program (see PC 2008).

Following the success of the ‘red books’ in encouraging GTE reform, with significant benefits for the Australian community, Australian governments recognised the potential to apply a similar performance reporting regime to government-provided services. These services not only accounted for a significant share of GDP, they were often provided to the most vulnerable members of the community. Even modest improvements in effectiveness and efficiency promised significant economic and social pay-offs. As the first report noted:

Improvements in the provision of these social services could benefit all Australians. The clients of the services could benefit by receiving services that are more relevant, responsive and effective. Governments could benefit by being encouraged to deliver the kinds of services that people want in a more cost effective manner. Taxpayers too could benefit from being able to see, for the first time in many cases, how much value they are receiving for their tax dollars, and whether services being provided effectively. (SCRCSSP 1995)

The creation of the Review in July 1993 established a systematic approach to reporting comparative data on the effectiveness and efficiency of government services. The original terms of reference are presented in box 9.1. These terms of reference were reaffirmed and extended by COAG in late 2009 (see attachment A).

Box 9.1 Key elements of original terms of reference

The Review, to be conducted by a joint Commonwealth/State and territory Government working party, is to undertake the following:

- Establish the collection and publication of data that will enable ongoing comparisons of the efficiency and effectiveness of Commonwealth and State government services ... this will involve:
 - establishing performance indicators for different services which would assist comparisons of efficiency and effectiveness. The measures should, to the maximum extent possible, focus on the cost effectiveness of service delivery, as distinct from policy considerations that determine the quality and level of services.
- Compile and assess service provision reforms that have been implemented or are under consideration by Commonwealth and State Governments.

From the outset, the RoGS embraced a diverse range of services, including education, health, justice, public housing and community services. The report also adopted a comprehensive approach to reporting on performance. In an era when most discussion of government services focused on the level of *inputs*, RoGS emphasised the importance of agreeing on the objectives of a service, and then creating robust indicators to measure the effectiveness, efficiency and equity of the services designed to achieve those objectives. Over time, the report has increasingly focused on the *outcomes* influenced by those services.

The 2011 RoGS report contained performance information for 14 ‘overarching’ service areas, encompassing 23 specific services (box 9.2).

RoGS’ coverage and scope have grown over time — the first report in 1995 addressed ten service sectors (*italicised* in box 9.2). Most recently, reporting on juvenile justice has been progressively introduced (as part of protection and support services) following a request from the Australasian Juvenile Justice Administrators. A mix of policy and pragmatism has guided the selection of service areas for reporting. Services are included that:

- make an important contribution to the community and/or economy (meaning there are potentially significant gains from improved effectiveness or efficiency)
- have key objectives that are common or similar across jurisdictions (lending themselves to comparative performance reporting)
- have relevant data collections, or data that could be collected relatively simply and inexpensively.

Box 9.2 Scope of RoGS 2011 vs RoGS 1995 (*italicised*)

Early childhood, education & training	<ul style="list-style-type: none">– Children’s services– School education<ul style="list-style-type: none">... <i>Government schools</i>... Non-government schools– <i>Vocational education and training</i>
Justice	<ul style="list-style-type: none">– Police services– <i>Court administration</i>– <i>Corrective services</i>
Emergency management	<ul style="list-style-type: none">– Fire, ambulance and road rescue services
Health	<ul style="list-style-type: none">– <i>Public hospitals</i>– Primary and community health– Breast cancer detection and management, and specialised mental health services
Community services	<ul style="list-style-type: none">– Aged care services– Services for people with disability– Protection and support services<ul style="list-style-type: none">... Child protection... Supported accommodation– Juvenile justice
Housing & homelessness services	<ul style="list-style-type: none">– Housing<ul style="list-style-type: none">... Public & community housing... Indigenous community housing... State owned and managed Indigenous housing... Commonwealth Rent Assistance– Homelessness services

Sources: SCRCSSP 1995; SCRGSP 2011.

Benchmarking and yardstick competition

The term ‘benchmarking’ can be used generally to refer to any process of comparison, but it also has a more technical meaning, implying specific steps and structured procedures designed to identify and replicate best practice (Vlasceanu *et al* 2004; Fenna, this volume).

RoGS does not establish benchmarks in the formal sense of systematically identifying best practice. Although some performance indicators are expressed in

terms of meeting particular standards (for example, measures of accreditation or clinically appropriate waiting times), most indicators have no explicit benchmark. That said, the information in the report can assist users to set their own benchmarks — in practice, the best jurisdiction’s performance, or the Australian average, may be treated as implied ‘benchmarks’.

There are sound reasons for RoGS’ focus on providing comparative information rather than formal benchmarking. From a policy perspective, it would be difficult for an inter-jurisdictional Steering Committee to come to a collective agreement on each other’s jurisdictions (see discussion below on the intergovernmental framework). More practically, the additional time required to analyse the large quantity of information contained in RoGS would significantly delay governments’ access to data needed in the budget cycle.

Further, any comparison of performance across jurisdictions requires detailed analysis of the potential impact of differences in clients, geography, available inputs and input prices. For example, a measure that shows relatively high unit costs in one jurisdiction may indicate inefficient performance, or may reflect better quality service, a higher proportion of special-needs clients or geographic dispersal. Across virtually all the services in the report, unit costs for the Northern Territory are significantly higher than for other jurisdictions, largely reflecting its relatively small and dispersed population, and high proportion of Indigenous Australians facing particular disadvantage. (That said, the Northern Territory still uses the report to compare other aspects of performance with the other jurisdictions, and to assess trends in unit costs over time).

To assist readers to interpret performance indicator results, the report provides information on some of the differences that might affect service delivery, including information for each jurisdiction on population size, composition and dispersion, family and household characteristics, and levels of income, education and employment. (Report content is discussed below). However, the report does not attempt to adjust reported results for such differences. Users of the report will often be better placed to make such judgments. As an aside, the methodology developed by the Commonwealth Grants Commission to allocate Commonwealth Government grants among the states and territories applies adjustment factors to account for the different costs of providing services in different jurisdictions (CGC 2010). These adjustment factors are contentious and subject to ongoing debate and refinement (see Banks, Fenna and McDonald this volume).

Productive vs unproductive competition

The maxim that ‘what gets measured gets managed’ is a particular issue when reporting on the provision of government services. The paucity of outcome and cost effectiveness indicators creates a risk that undue emphasis will be placed on necessarily partial input and output indicators. As competitive pressures mount (for example, where financial rewards or penalties are based on reported performance) so do the risks of goal displacement (chasing the proxy measure, rather than the desired outcome), or manipulation of data (see Fenna, this volume).

From the outset, the Steering Committee responsible for the RoGS has sought to manage such risks. The structure of the Review of Government Service Provision (see discussion below on governance arrangements) involves a consultative approach to identifying service objectives and indicators, ensuring reporting is appropriate and balanced. The RoGS performance indicator framework emphasises the importance of considering all aspects of performance and explicitly identifies any significant gaps in reporting. To encourage readers to seek out indicator detail (including data caveats and relevant and context), the Steering Committee has resisted summary ‘traffic light’ or ‘dashboard’ approaches to presentation.

Finally, the Steering Committee places considerable weight on reporting high quality data. Reporting aligns with nationally agreed data definitions and draws on data collected and verified by national statistical agencies wherever possible. At a minimum, all data have been endorsed by the contributor and subjected to peer review by a working group made up of representatives of relevant line agencies from all jurisdictions.

Synergies with other national reporting exercises

A number of the services included in RoGS are subject to other performance measurement exercises, typically at a sectoral level. For example, relevant Ministerial Councils commission annual national reports on schools and hospitals (MCEEDYA 2008; AIHW 2010). It would be a concern if RoGS merely duplicated information reported elsewhere (although once data are collected, the marginal cost of reproducing them in different reports for different purposes or different audiences is minimal.) However, RoGS has several features that distinguish it from other reports.

First, a Steering Committee of senior officials from central agencies sets it apart from most other national reporting exercises, which are driven by line agencies or data agencies. The content and approach of RoGS have been strongly influenced by the Steering Committee’s priorities, for example:

-
- *making use of available data* — data are reported for those jurisdictions that can (or are willing) to report, rather than waiting for completeness or unanimity. Experience has shown that once a few jurisdictions report, other jurisdictions soon follow suit
 - *no jurisdictional veto* — a jurisdiction can withhold its own data from publication but cannot veto the publication of another jurisdiction's data (unlike some Ministerial Council publications)
 - *providing policy makers with timely data* — even where there may be a trade-off with data quality. The following general test is applied: 'are policy makers better off with these data (even qualified) than no data at all'. Of course, data that are likely to mislead are not reported, and imperfect data are caveated in the report. Publication increases scrutiny of the data and tends to encourage improvement in data quality over time
 - *producing an accessible report* — the report is aimed at a non-technical audience. Indicators are designed to be intuitive and non-ambiguous, and explained in lay terms.

Second, RoGS reports on the various service areas according to a consistent, structured framework in a single, annual report (see below). In addition to providing a convenient resource for people interested in more than one service area, this approach has strategic and practical benefits. Strategically, experience has shown that jurisdictional 'winners' and 'losers' tend to vary across the reported services, making it easier for Steering Committee members to 'hold the line' on reporting in those areas where their jurisdiction performs relatively poorly. More pragmatically, having working groups and data providers working to the same timetable creates 'positive pressure' for both timeliness and continuous improvement.

Third, unlike many sectoral reports, RoGS explicitly addresses all dimensions of performance — equity and efficiency, as well as effectiveness. Data are gathered from a range of sources for each service area, to ensure all dimensions are covered (including Secretariat collections to address data gaps). Often, data are recast into agreed performance indicators, involving the transformation or further disaggregation of data published elsewhere. As noted, the report also identifies any gaps in reporting, alerting readers to aspects of performance not currently measured, and placing pressure on departments and data agencies to improve data collection.

9.2 The intergovernmental framework

As noted, RoGS' original mandate came from an explicit agreement of heads of government in 1993 (box 9.1). In December 2009, following a high level review,

COAG (2009) agreed that RoGS should continue to be the key tool to measure and report on the efficiency and effectiveness of government services. COAG endorsed new, expanded, terms of reference for the Steering Committee and the RoGS, and a charter of operations formalising many of the existing Steering Committee operating principles (attachments A and B respectively). COAG also noted the complementary role of the COAG Reform Council, analysing and reporting on National Agreement outcomes and performance benchmarks (see Banks, Fenna and McDonald, this volume; O’Loughlin, this volume).

Purpose and audience

As the terms of reference make clear, RoGS is primarily a tool for government — although the 2009 review confirmed public accountability as an important secondary purpose.

Performance measurement can promote better outcomes, first by helping to clarify government objectives and responsibilities, and then by making performance more transparent, enabling assessment of whether and how well program objectives are being met. Well-structured performance measurement, with a comprehensive framework of indicators, provides a means of widening the focus from resourcing to the efficient and effective use of those resources. It can also encourage analysis of the relationships between programs, assisting governments to coordinate policy within and across agencies.

Comparative performance reporting offers three additional advantages. It allows governments, agencies and clients to verify high performance. The identification of successful agencies and service areas provides opportunities for governments and agencies to learn from counterparts delivering higher quality or more cost-effective services. And ‘yardstick competition’ can generate pressure for improved performance (see Fenna, this volume).

Surveys of users of the report have identified that RoGS is used for strategic budget and policy planning, and for policy evaluation. Information in the report has also been used to assess the resource needs and resource performance of departments. And it has been used to identify jurisdictions with whom to share information on services (SCRGSP 2007).

Governance arrangements have been pivotal

The Review's governance arrangements drew on the innovative model developed for the GTE (red book) process, and have played a key role in the success of the RoGS. Two particular design features have been instrumental:

- the combination of top-down policy with bottom-up expertise
- the independence of the Steering Committee's chairman and secretariat.

Top-down policy, bottom-up expertise

The first key design feature is the combination of 'top-down' authority exercised by a Steering Committee of senior officials from central agencies, with 'bottom up' expertise contributed by line agency working groups.

The Steering Committee comprises senior representatives from the departments of first ministers, and treasury and finance. It provides high-level strategic direction, as well as the authority and drive required to encourage services to report transparently on performance. There have been many instances where the Steering Committee's whole-of-government perspective has been crucial in resisting the short term imperatives that can, at times, dominate line agency priorities.

The Steering Committee has often been a 'first mover' in identifying gaps in reporting and pressing for the development of related performance indicators. The Steering Committee pioneered the inclusion of data on the user cost of capital in financial data reporting, for example, and was instrumental in encouraging the introduction of nationally comparable learning outcomes. The Steering Committee has also ensured that important indicators continue to be reported despite occasional reluctance from line agencies (for example, elective surgery waiting times by urgency category and court administration backlogs).

Working groups comprise senior line agency experts. They provide necessary subject area expertise, and ensure the report is grounded in reality. Cross membership of working groups and related parallel groups (such as Ministerial Council committees and COAG working groups) has helped RoGS to remain aligned with governments' strategic and policy priorities.

An independent chair and secretariat

The second key design feature is the independence of the principal governance arrangements. Although a Commonwealth Government authority, the Productivity Commission operates under a statute that enables it to act independently of the

interests of any jurisdiction, portfolio or data provider. In its work, the Commission has acquired a reputation for impartiality and transparency, as well as for rigorous analysis directed at enhancing the interests of the community as a whole.

The Commission's 'honest broker' status helped neutralise early concerns that the exercise would be dominated by the Commonwealth government and imposed upon State governments. Having an impartial Chairman and secretariat has helped foster a collaborative and cooperative environment, and facilitated consensus decision-making on potentially contentious issues.

Over time, the work of the Steering Committee and its secretariat has expanded to produce other reports for COAG, including the *Overcoming Indigenous Disadvantage* report, National Agreement performance reporting and the *Indigenous Expenditure* Report, creating useful synergies. (All Steering Committee reports are available from the Review website at: www.pc.gov.au/gsp). The broader inquiry and research work of the Productivity Commission in turn has benefited from the Secretariat's performance reporting expertise.

9.3 The RoGS approach to reporting

Report content

The main focus of RoGS is information on comparative performance, but RoGS also provides a range of additional material to assist users to interpret the performance data. The report includes introductory chapters that explain the approach to performance reporting and recent developments in the report.

A sector preface introduces each set of related chapters (that is, 'early childhood, education and training'; 'justice'; 'health'; 'community services' and 'housing and homelessness'). Each preface provides an overview of the sector and any cross-cutting or interface issues, and reports some high level performance information.

Each chapter provides a profile of the relevant service area, including a discussion of the roles and responsibilities of each level of government, and a statement of the agreed service objectives. Some general descriptive statistics about the service area are provided as context. Each chapter also includes one page for each jurisdiction to comment on their reported performance or highlight policy and program initiatives. This has provided a useful 'safety valve', allowing jurisdictions to provide their own interpretation of reported results, or steps being taken to improve performance, in circumstances where they may otherwise have withdrawn their data.

A statistical appendix provides further information to assist the interpretation of the performance indicators presented in the report, including information on each jurisdiction’s population, family and household characteristics, income, education and employment, and explanations of statistical concepts used in the report.

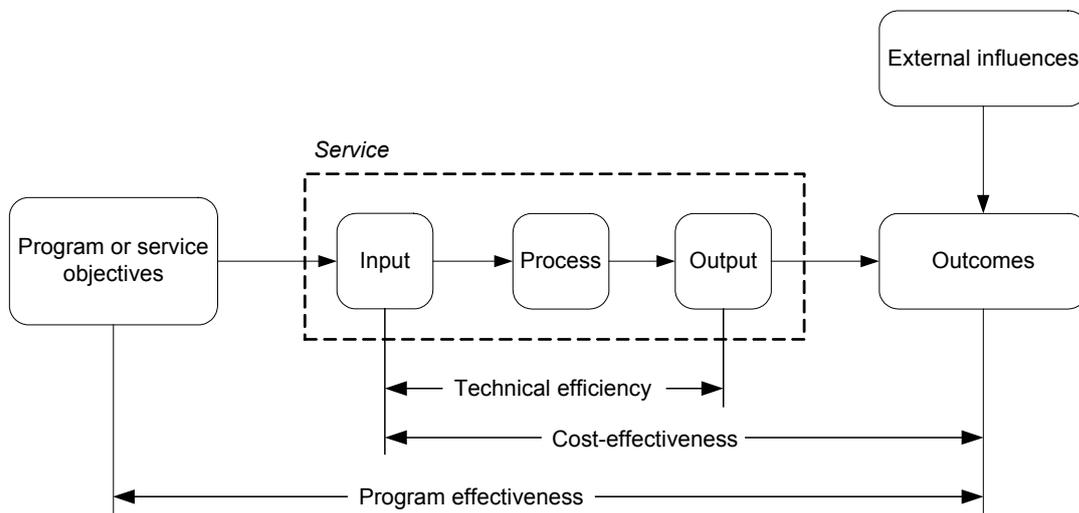
The performance indicator framework

The Steering Committee has developed a generic performance indicator framework that is applied to all services areas in RoGS, although individual service areas may tailor the framework to reflect their specific objectives or to align with other national reporting frameworks.

The RoGS general framework reflects the ‘service process’ by which service providers transform inputs (resources) into outputs (services), in order to achieve agreed objectives. Figure 9.2 identifies the following aspects of the service process:

- program effectiveness (the achievement of objectives)
- technical efficiency (the rate of conversion of inputs to outputs)
- outcomes (the impact of services on individuals or the community).

Figure 9.2 **Service process**



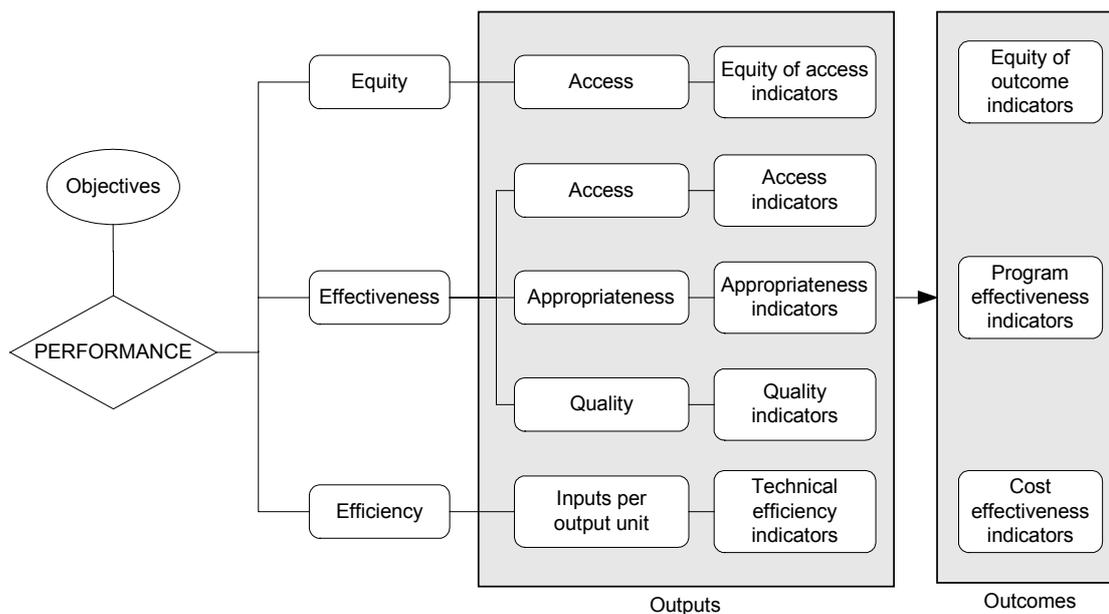
Source: SCRGSP (2011), p. 1.14.

The indicator framework

The indicator framework has evolved over time. The current general performance framework is set out in figure 9.3.

The original framework was based on effectiveness and efficiency; it did not separately identify equity, or clearly distinguish outputs and outcomes. The current framework highlights the importance of outcomes, even though these are typically difficult to measure. It is also difficult to isolate the specific impact of government services, given other influences outside the control of service providers (Fenna, this volume). The Steering Committee acknowledges that services provided by government may be only one contributing factor to outcomes and, where possible, RoGS includes information on other factors, including different geographic and demographic characteristics across jurisdictions. The performance indicator framework therefore includes information on outputs — the services actually produced — as proxies for outcome measures, where evidence suggests a direct link between those outputs and the objectives of the service. Output information is also necessary to inform the management of government services, and is often the level of performance information of most interest to service users.

Figure 9.3 **General performance indicator framework**



Source: SCRGSP (2011), p. 1.13.

The indicator framework groups output indicators according to the desired characteristics of a service, including:

- *Efficiency indicators* — measures of how well organisations use their resources, typically being measures of technical efficiency (that is, (government) inputs per unit of output).
- *Effectiveness indicators* — measures of whether services have the sorts of characteristics shown to lead to desired outcomes:

-
- access (availability and take-up of services by the target population)
 - appropriateness (delivery of the right service)
 - quality (services that are fit for purpose, or measures of client satisfaction).
 - *Equity indicators* — measures of access for identified ‘special needs groups’, including Indigenous Australians, people with disability, people from culturally diverse backgrounds, people from regional and remote locations and, depending on the service, particular sexes or age groups.

An ‘interpretation box’ for each indicator provides the definition of the indicator measure, advice on interpretation of the indicator, any data limitations, whether the reported measures are complete and/or fully comparable. Where data are not directly comparable, appropriate qualifying commentary is provided in the text or footnotes. Where data cannot be compared across jurisdictions, time series data allows the assessment of a jurisdiction’s performance over time.

Cross-cutting and interface issues

Governments are increasingly focused on achieving outcomes that involve more than one service area. For example, increases in the proportion of older people in the population are raising demand for aged care and disability services, with an emphasis on coordinated community services that limit the need for entry into institutional care. Similarly, access to effective community services may influence outcomes for clients of education, health, housing and justice sector services.

Although these issues are difficult to address in a report structured by service area, the Steering Committee has tried to break down the service-specific ‘silos’ through innovations such as a ‘health management’ chapter (which reports on management of diseases, illnesses and injuries using a range of services (promotion, prevention/early detection and intervention) in a variety of settings (for example, public hospitals, community health centres and general practice). It has also enhanced section prefaces with high-level measures of sector-wide performance, and provided extensive cross-referencing throughout the report.

Production processes

Publication

RoGS currently consists of an annual, two-volume hard copy publication containing the chapters, prefaces and appendix, supported by electronic data attachments

available through the Review website. (The chapters, prefaces and appendix are also available electronically). The Steering Committee has considered moving to solely electronic publication but key users prefer receiving hard copies.

Timetable

The current publication date, at the end of January each year, was agreed by the Steering Committee to maximise the potential for RoGS to inform the annual budget cycle. To meet the publication date, working groups and the Secretariat follow the timetable outlined in box 9.3. Jurisdictions comment on two drafts of the report before sign-off.

Box 9.3 Report on Government Services timetable

- March — working groups agree on strategic plans for next (and future) reports
- April — Steering Committee endorses strategic plans
- June/July — working groups agree on content of next report
- End-July — Secretariat finalises data manuals and circulates data requests
- August — Steering Committee agrees on developments for next report
- End-September — Data deadline (subject to agreed extensions)
- End-October — Secretariat circulates working group draft
- November — working groups comment on working group draft
- End-November — Secretariat circulates Steering Committee draft
- Early-December — Steering Committee comments on Steering Committee draft
- Mid-December — Secretariat circulates final draft for sign off out of session
- January — Secretariat finalises report and manages printing and distribution
- End-January — Report published

Data management

Data for RoGS are collected from some 200 data providers, largely using Excel spreadsheets. These data are then stored and manipulated using a customised database, developed for the 2004 RoGS. With recent improvements in information technologies there is scope to modernise RoGS data collection, manipulation and reporting, although this would require a significant one-off investment in updating systems.

Costs versus benefits?

The costs of producing the RoGS are significant. They include not only the Secretariat's costs (approximately \$2.8 million, mostly for staff), but also those of government agencies (19 Steering Committee members and 180 working group members) and over 200 data providers. Although a formal cost–benefit analysis has not been undertaken, there is plenty of circumstantial evidence that the information in RoGS has played a significant role in informing policy improvements across a broad range of services. Given the economic and social importance of the services covered by RoGS, even relatively small improvements in their effectiveness or efficiency would be expected to far outweigh the cost of producing it.

9.4 Conclusions and some lessons

How successful?

Looking back over its 15-year history, the review could lay claim to being one of the success stories of cooperative federalism in Australia. It has proven an effective vehicle for delivering agreement across governments about what matters for performance, and for the collection and publication of robust data to inform performance comparisons. This achievement has been remarkable on a number of fronts — not least the ongoing commitment of heads of government to the production of what is effectively an annual 'report card' on their performance across an array of politically sensitive services.

The fact that RoGS does not include overt analysis or recommendations makes it difficult to draw direct links to specific policy or program reforms. However, there is extensive circumstantial evidence that the information in RoGS has played a significant role in informing policy development across a broad range of services. To take some examples:

- In the education sector, the Steering Committee was instrumental in the introduction of standardised national testing of student learning outcomes, the results of which are now galvanising education departments around Australia.
- In the health sector, RoGS reporting illustrated the beneficial impact of the introduction of 'case mix' funding by Victoria on the average cost of hospital separations. Over time, other jurisdictions introduced some form of activity based costing of hospital services, and the approach is now being adopted at a national level.

-
- In the justice sector, RoGS reporting illustrated the significant efficiency gains associated with Victoria's use of electronic courts for minor traffic infringements, which soon spread to other jurisdictions.
 - In the community services sector, the Steering Committee was instrumental in developing and reporting Indigenous 'potential populations' for disability services, demonstrating that the previous unadjusted population rates significantly overstated Indigenous peoples' access to services relative to their level of need.
 - In the housing sector, development and reporting of comparable data for mainstream and Indigenous-specific social housing (an ongoing task) has highlighted the potential for differential standards for essentially similar services.

Channels of influence

RoGS appears to have influenced policy and encouraged improvements in government service delivery through four broad mechanisms or channels.³

First, governments have benefited simply from having to respond to the information requirements of the RoGS process. Particularly in its early years, RoGS drove significant improvements in basic management information. In order to provide data to RoGS, many services had to upgrade their rudimentary information systems.

The Steering Committee's reporting framework also forced all jurisdictions to clarify and agree on the objectives of each government service, and to define how 'success' would be measured. This was a challenge for many service sectors, with sometimes-heated debates over the appropriate role of government; for example, whether the objective of children's services was to facilitate parents' labour market participation, or to promote the development of children.

Steering Committee and working group meetings also provide regular opportunities for the informal sharing of information. Members share experiences of reforms and assist each other to improve data and its analysis. Members have often gone on to collaborate outside formal RoGS processes, to the mutual benefit of their jurisdictions.

A second, related source of benefit has been the opportunity for each government to learn more about their *own* jurisdiction. Steering Committee members report that

³ More information and specific examples can be found in chapter 2 and appendix B of the Productivity Commission's Annual report (PC 2010).

peer pressure through the Review often has aided them in extracting information from line agencies that previously had not been obtainable. More generally, the Steering Committee/working group structure has contributed to better two-way understanding, within each government, of both central agency strategic priorities and line agency constraints and capabilities.

Third, governments and citizens have benefited from what they have learnt about the performance of other jurisdictions. Typically, ministers and senior executives in all jurisdictions are briefed on the performance of their portfolios and agencies before the release of each report. Service areas are often required to justify perceived ‘underperformance’ relative to their counterparts in other jurisdictions. Further, comparative data from RoGS are cited extensively within Australia’s eight parliaments and in parliamentary committees; are drawn on in performance audits by the federal and State audit offices; and are cited in policy review documents

Fourth, RoGS has become a key accountability tool and a resource that is also utilised outside government. Each year, the report receives extensive media coverage, disseminating its information to a wide audience. This in turn tends to generate public pressure for governments to justify perceived poor performance, and to improve performance over time. The iterative nature of RoGS has contributed to better understanding of the information by the media and improvements in responsible (or ‘accurate’) reporting over time.

Information in RoGS is also drawn on by many community groups, both for advocacy purposes, and as a tool for assessing their own performance where they deliver services on behalf of governments. (The Steering Committee has recently endorsed a proposal from Monash university academics to partner with the Secretariat to investigate the use of RoGS by the non-government sector, initially focusing on members of the Victorian Council of Social Services.) There is also widespread use of RoGS by government researchers, university academics and consultants, across a wide range of disciplines.

Some key contributors to this success

This paper has already identified several aspects of the ‘design’ and operation of the Review that have contributed to its effectiveness and longevity. The most notable are a governance structure that allows the strategic direction of the review to be set by senior officials of central agencies, with the benefit of line agency expertise; and a chair and secretariat that are independent of the interests of any jurisdiction, portfolio or data provider.

The Review has also benefited from the close involvement of Australia's national data agencies, the Australian Bureau of Statistics and the Australian Institute of Health and Welfare. Comparative performance reporting for many services was facilitated by the existence of mandatory National Minimum Data Sets, established as part of the system of financial transfers between the Commonwealth Government and the states and territories (see Banks, Fenna and McDonald, this volume). That said, there are still many significant data gaps, and a need for governments to fund the evidence base that we need to compare performances across the federation (Banks 2009). An important recent initiative in this direction is the allocation of additional funding for a new performance reporting framework for schools and hospitals (the 'MySchool' and 'MyHospital' programs). The then Minister for Education (now Prime Minister), Julia Gillard, noted in endorsing the new schools framework:

It is my strong view, that lack of transparency both hides failure and helps us ignore it...And lack of transparency prevents us from identifying where greater effort and investment are needed' (Gillard 2008).

Another factor has been the development of a performance indicator framework based on a 'service process' model. Reporting consistently across a wide range of services in a single report has facilitated the cross-fertilisation of ideas, and made it easier for Steering Committee members to 'hold the line' in areas where their jurisdiction's service performance looks relatively poor. It has also created peer pressure to maintain timeliness and improve reporting.

Room for improvement

With its ethos of performance improvement, the Review is acutely aware of the need for continuous improvement in its own work. The Steering Committee, working groups and Secretariat undertake an annual strategic planning process to evaluate their own performance and identify scope to enhance processes and report content. The Steering Committee regularly surveys report users as to their satisfaction with RoGS and ideas for improvement (SCRGSP 2007).

Most recently, the Steering Committee has benefited from the findings of the 2009 review of RoGS (COAG 2009), which, among other things, recommended new terms of reference (see attachment A).

Highlighting improvement and innovation in service delivery

As noted, there is circumstantial evidence that the comparative data in RoGS help drive improvements in service delivery. However, the links between those data and reforms to service delivery can be indirect, and are rarely acknowledged publicly.

Governments have been seeking a mechanism by which comparative performance reporting can drive reform more directly. The review of RoGS recommended that the Steering Committee should highlight improvements and innovations in service delivery by selecting a small number of subjects to be developed as case studies — what Fenna (this volume) describes as the ‘qualitative dimension of benchmarking’. This reinforces an aspect of the original terms of reference — ‘to compile and assess service provision reforms’ — that lost impetus after an initial burst of enthusiasm. The Steering Committee has agreed to include ‘mini case studies’ in RoGS, and to consider undertaking more substantial research into improvements and innovations in service delivery.

Some final comments

The competitive and cooperative dimensions of Australia’s federal system both have roles to play in helping address the significant policy challenges that lie ahead, including population ageing and increasing demands for more, and better quality, health, education and community services.

The RoGS has proven an effective and enduring mechanism for harnessing these competitive and cooperative dimensions to benefit Australia’s community.

Notwithstanding many improvements over the years, there is considerable scope for further reform in government service provision. The Productivity Commission’s report for COAG on the benefits of the National Reform Agenda suggests that reforms in human services and other policy areas bearing on human capital development could yield gains as substantial as those from earlier, competition-related reforms (PC 2006). The publication of comparable performance data across Australia’s jurisdictions has a significant role to play in facilitating those reforms.

Attachment A Terms of reference

Steering Committee terms of reference

- | | |
|---|--|
| <p>(1) The Steering Committee for the Review of Government Service Provision (the Steering Committee) was established by the Council of Australian Governments (COAG) and comprises representatives of the Commonwealth, State and Territory governments.</p> | Constitution and authority of Steering Committee |
| <p>(2) The Steering Committee will operate according to a Charter of Operations.</p> | |
| <p>(3) As an integral part of the national performance reporting system, the Steering Committee informs Australians about services provided by governments and enables performance comparisons and benchmarking between jurisdictions and within a jurisdiction over time. The Steering Committee and its working groups are supported by a Secretariat located within the Productivity Commission as a neutral body that does not represent any jurisdiction.</p> | Objectives |
| <p>(4) Better information improves government accountability and contributes to the wellbeing of all Australians by driving better government service delivery. To this end, the Steering Committee will:</p> <ul style="list-style-type: none">i. measure and publish annually data on the equity, efficiency and cost effectiveness of government services through the Report on Government Servicesii. produce and publish biennially the Overcoming Indigenous Disadvantage reportiii. collate and prepare performance data under the Intergovernmental Agreement on Federal Financial Relations, in support of the analytical role of the COAG Reform Council and the broader national performance reporting systemiv. initiate research and report annually on improvements and innovation in service provision, having regard to the COAG Reform Council's task of highlighting examples of good practice and performance perform any other related tasks referred to it by COAG. | Outputs |
| <p>(5) The Report on Government Services and the Overcoming Indigenous Disadvantage report will be produced subject to additional terms of reference.</p> | |

Continued next page

Steering Committee terms of reference (continued)

- (6) To support the quality and integrity of these products, the Steering Committee will:
- i. ensure the integrity of the performance data it collects and holds
 - ii. exercise stewardship over the data, in part through participation in data and indicator development work of other groups that develop, prepare and maintain data used in Review reports, and through reporting outcomes of Steering Committee data reviews to authorities such as Heads of Treasuries and COAG, to ensure its long term value for comparisons of government service delivery, and as a research and evidence tool for the development of reforms in government service delivery
 - iii. ensure that performance indicators are meaningful, understandable, timely, comparable, administratively simple, cost effective, accurate and hierarchical, consistent with the principles for performance indicators set out under the Intergovernmental Agreement on Federal Financial Relations
 - iv. keep abreast of national and international developments in performance management, including the measurement and reporting of government service provision.
- (7) The Steering Committee's ability to produce meaningful comparative information requires timely access to data and information. All jurisdictions have committed to facilitate the provision of necessary data, either directly or via a data agency, to meet Steering Committee timelines and to ensure the Steering Committee can meet its obligations to COAG.
- (8) The Steering Committee will seek to maximise the accessibility to governments and the Australian community of the performance data it collects and collates, taking advantage, where appropriate, of developments in electronic storage, manipulation and publication of data. It will work with other government agencies in Australia undertaking similar work to ensure a consistent and best practice approach.

Data quality and integrity

Accessibility

Continued next page

Steering Committee terms of reference (continued)

(9) The Steering Committee will also, subject to direction from COAG, and in recognition of its role in the broader national performance reporting framework:

Relationships
within the national
performance
reporting system

- i. have regard to the work program of the COAG Reform Council and provide such data as is required by the Council for the performance of its functions
- ii. align, insofar as possible, the data collected and indicators developed with those under the National Agreements, avoiding duplication and unnecessary data collection burdens on jurisdictions
- iii. drive improvements in data quality over time, in association with the Ministerial Council for Federal Financial Relations, the COAG Reform Council, other Ministerial Councils and data agencies.

Source: COAG 2010.

Report on Government Services terms of reference

- | | |
|--|-------------------------------------|
| <p>(1) The Steering Committee will measure and publish annually data on the equity, efficiency and cost effectiveness of government services through the Report on Government Services (RoGS).</p> | <p>Outputs and objectives</p> |
| <p>(2) The RoGS facilitates improved service delivery, efficiency and performance, and accountability to governments and the public by providing a repository of meaningful, balanced, credible, comparative information on the provision of government services, capturing qualitative as well as quantitative change. The Steering Committee will seek to ensure that the performance indicators are administratively simple and cost effective.</p> | |
| <p>(3) The RoGS should include a robust set of performance indicators, consistent with the principles set out in the Intergovernmental Agreement on Federal Financial Relations; and an emphasis on longitudinal reporting, subject to a program of continual improvement in reporting.</p> | |
| <p>(4) To encourage improvements in service delivery and effectiveness, RoGS should also highlight improvements and innovation.</p> | |
| <p>(5) The Steering Committee exercises overall authority within the RoGS reporting process, including determining the coverage of its reporting and the specific performance indicators that will be published, taking into account the scope of National Agreement reporting and avoiding unnecessary data provision burdens for jurisdictions.</p> | <p>Steering Committee authority</p> |
| <p>(6) The Steering Committee will implement a program of review and continuous improvement that will allow for changes to the scope of the RoGS over time, including reporting on new service areas and significant service delivery areas that are jurisdiction-specific.</p> | |
| <p>(7) The Steering Committee will review the RoGS every three years and advise COAG on jurisdictions' compliance with data provision requirements and of potential improvements in data collection. It may also report on other matters, for example, RoGS's scope, relevance and usefulness; and other matters consistent with the Steering Committee's terms of reference and charter of operations.</p> | <p>Reporting to COAG</p> |

Source: COAG 2010.

Attachment B Charter of operations

Review of Government Services charter of operations

- | | |
|--|-------------|
| <p>(1) This charter of operations sets out the governance arrangements and decision making processes for the Steering Committee for the Review of Government Service Provision (the Steering Committee). It should be read in conjunction with the Council of Australian Governments (COAG)-endorsed terms of reference for the Steering Committee. Additional information on the Steering Committee's policies and principles can be found in the introductory chapters of relevant reports and the 'Roles and responsibilities of Review participants' document.</p> | Preamble |
| <p>(2) COAG established the Steering Committee in 1993, to produce ongoing comparisons of the efficiency and effectiveness of Commonwealth, State and Territory government services (through the Report on Government Services [RoGS]) and to compile and assess service provision reforms.</p> | History |
| <p>(3) In December 2009, COAG confirmed the RoGS should continue to be the key tool to measure and report on the productive efficiency and cost effectiveness of government services, as part of the national performance reporting system.</p> | |
| <p>(4) The Steering Committee comprises senior officials from the central agencies (First Ministers, Treasuries and Finance departments) of the Commonwealth, States and Territories. The Steering Committee is chaired by the Chairman of the Productivity Commission.</p> | Membership |
| <p>(5) In recognition of the value of expert technical advice, and the need for collaborative action, the Steering Committee may include observers from relevant data agencies.</p> | Observers |
| <p>(6) The Steering Committee and its working groups are supported by a Secretariat located within the Productivity Commission. The Secretariat is a neutral body and does not represent any jurisdiction.</p> | Secretariat |
| <p>(7) The Steering Committee may establish working groups, cross-jurisdictional or otherwise, to provide expert advice. Working groups typically comprise a convenor drawn from the membership of the Steering Committee and State, Territory and Commonwealth government representatives from relevant departments or agencies. Working group members should have appropriate seniority to commit their jurisdictions on working group matters and provide strategic policy advice to the Steering Committee.</p> | |

(Continued next page)

Review of Government Services charter of operations (continued)

- (8) In recognition of the value of expert technical advice and close relationships with data development bodies and agencies, working groups may include observers from relevant data agencies or, where a data agency is not available, Ministerial Council data sub-committees. Furthermore, working groups may consult with data agencies or sub-committees, as appropriate, on technical issues requiring expert consideration. Working groups
- (9) Working groups may contribute to and comment on drafts of Steering Committee reports, and make recommendations to the Steering Committee on matters related to their areas of expertise.
- (10) Working groups are advisory bodies and do not endorse report content. As far as practicable, working groups adopt a consensus approach to making recommendations to the Steering Committee. Where working groups do not reach consensus, alternative views should be provided to the Steering Committee for decision.
- (11) As far as practicable, the Steering Committee adopts a consensus approach to decision-making. Where consensus is not reached, decisions are based on majority vote of Steering Committee members, with each jurisdiction's members having one joint vote. (Observers may not vote.) Should the Steering Committee be equally divided, the Chairman has a casting vote. Governance and decision-making arrangements
- (12) Steering Committee members from one jurisdiction may choose not to publish information relating to their own jurisdiction but may not veto the publication of information relating to other jurisdictions.
- (13) The Steering Committee may draw on the expert advice of its Secretariat, working groups and of specialist data and other organisations, but it is not bound by such advice.

Source: COAG 2010.

References

- AIHW (Australian Institute of Health and Welfare) 2010, *Australian Hospital Statistics*, Cat. no HSE 84, AIHW, Canberra.
- Banks, G. 2002, *Reviewing the Service Performance of Australian Governments*, (aAddress to International Quality & Productivity Centre summit 'Measuring and Managing Government Performance', 20 February 2002, Canberra), Productivity Commission, Canberra.
- 2005, *Comparing school systems across Australia* (Address to ANZSOG conference, 'Schooling in the 21st Century: Unlocking Human Potential', 28 and 29 September 2005, Sydney), Productivity Commission, Canberra.
- 2009, *Evidence-based policy making: What is it? How do we get it?* (ANU Public Lecture Series, presented by ANZSOG, 4 February), Productivity Commission, Canberra.
- COAG (Council of Australian Governments) 2009, *COAG Communiqué 7 December 2009*, www.coag.gov.au/coag_meeting_outcomes/2009-12-07/docs/agenda_item_report_government_services_review_executive_summary.rtf.
- Commonwealth Grants Commission 2010, *Report on GST Revenue Sharing Relativities — 2010 Review*, Canberra.
- Gillard, J. (Deputy Prime Minister) 2008, *Leading transformational change in schools*, Address to the Leading Transformational Change in Schools Forum, Melbourne, 24 November.
- MCEEDYA (Ministerial Council for Education, Early Childhood Development and Youth Affairs) 2008, *National Report on Schooling in Australia*, <http://cms.curriculum.edu.au/anr2008/index.htm>.
- PC (Productivity Commission) 2006, *Potential Benefits of the National Reform Agenda*, Report to the Council of Australian Governments, Canberra.
- 2008, *Financial Performance of Government Trading Enterprises, 2004-05 to 2006-07*, Commission Research Paper, Canberra, July.
- 2009, *Submission to the House of Representatives Standing Committee on Economics: Inquiry into Raising the Level of Productivity Growth in Australia*, September.
- 2010, *Annual Report*, Canberra.
- 2011, *Annual Report*, Canberra.
- SCNPMGTE (Steering Committee on National Performance Monitoring of Government trading Enterprises) 1992, *Measuring the Total Factor Productivity*

of Government Trading Enterprises, Performance Monitoring Report, Industry Commission [now Productivity Commission], Canberra.

SCRCSSP (Steering Committee for the Review of Commonwealth/State Service Provision) 1995, *Report on Government Services 1995*, Industry Commission [now Productivity Commission], Canberra.

SCRGSP (Steering Committee for the Review of Government Service Provision) 2007, *Feedback on the Report on Government Services 2007*, Productivity Commission, Canberra.

— 2011, *Report on Government Services 2011*, Productivity Commission, Canberra.

Vlăsceanu, L., Grünberg, L., and Pârlea, D. 2004, *Quality Assurance and Accreditation: A Glossary of Basic Terms and Definitions* (Bucharest, UNESCO-CEPES) Papers on Higher Education, ISBN 92-9069-178-6. <http://www.cepes.ro/publications/Default.htm>.

10 Benchmarking, competitive federalism and devolution: how the COAG reform agenda will lead to better services

Ben Rimmer¹

Department of Human Services, Australian Government

10.1 Introduction

Since December 2007, the Council of Australian Governments (COAG) has embarked on a major new agenda of national reform — the ‘COAG reform agenda’. The objective of the COAG reform agenda is to improve the well-being of all Australians. The COAG Reform Council describes this as the ‘most comprehensive economic, social and environmental reform agenda ever contemplated in the context of intergovernmental relations in Australia’ (CRC 2010a, p. xii).

This paper argues that the COAG reform agenda reflects three elements that are critical to improved service delivery:

1. funding linked to the achievement of outcomes and outputs (rather than inputs) in areas of policy collaboration
2. devolution of decision making and service design to the frontline, and
3. competitive tensions between the States and Territories (‘competitive federalism’) and competitive tensions between service providers.

Underpinning these elements is a cornerstone of the COAG reform agenda: increased transparency and the use of benchmarking to measure performance.

¹ Ben Rimmer is Associate Secretary, Service Delivery, in the Australian Government Department of Human Services. He was Deputy Secretary, Strategic Policy and Implementation, Department of the Prime Minister and Cabinet from 2008–2011. The views expressed in this paper are the author’s own and do not necessarily reflect the official views of either department or the Australian Government. The author acknowledges the assistance of Toby Robinson in the Department of the Prime Minister and Cabinet in developing this paper.

The COAG reform agenda is an example of a broader trend in policy design, particularly in service delivery, which involves ‘market design’. In other words, the COAG reform agenda moves beyond the outdated and bifurcated debate about the merits of wholly public sector or wholly private sector service delivery. It instead focuses on how governments can design interventions in markets to achieve policy outcomes. Each of the elements listed above is fundamental to this process, and transparency and the use of benchmarking are at its heart.

Remodelling the Commonwealth–State relationship

This paper also argues that the COAG reform agenda can be viewed as a modern remodelling of the relationship between the Commonwealth and the States and Territories (‘the States’). As part of that remodelling, the Commonwealth offered the States increased funding; interventions that were better targeted on specific COAG-agreed reforms (such as through National Partnership Agreements); devolution and flexibility in decision making and service design; and a more engaged and collaborative approach to the task of national policy leadership. In return, the Commonwealth sought from the States much greater levels of transparency; more innovation and responsiveness in policy development and service delivery; better use of Commonwealth funding; and assurances that the States would follow through on COAG-agreed reforms and ensure delivery of shared national objectives.

The COAG reform agenda has already delivered significant benefits. This paper explores some of the institutional successes, such as the strengthened role of the COAG Reform Council (CRC). It also explores some of the policy successes, including examples from the significant reform effort now underway that is leading to better services for the Australian community. These successes help to illustrate how, on the whole, the Australian federation works well, despite occasional hiccups.

Challenges and risks

Despite the successes, the jury is still out on whether the full potential of the COAG reform agenda is being delivered. There are also some clear risks to the COAG reform agenda. Some policy debates, such as health reform, have explicitly moved to a related but separate institutional framework. In a period of fiscal consolidation, it will be difficult for the Commonwealth to preserve current funding levels for a number of initiatives.

Unless some of these risks dissipate, the Commonwealth will find it difficult to continue to offer the States the increased levels of flexibility in policy and service delivery design that the COAG reform agenda has provided. Under pressure for faster and better delivery, there are a range of alternative proposals for reform in key policy areas that the Commonwealth could adopt, which would offer much less flexibility to the States but might be seen by Commonwealth Ministers to deliver more to the nation.

For believers in Australian federalism, now is the time for delivery. Barring major reallocations of roles and responsibilities within the federation, achieving nationally significant reforms will continue to require collaboration between the Commonwealth and the States. Funding linked to outcomes and outputs, greater devolution, competitive federalism and, importantly, increased transparency and benchmarking, will all remain critical to improving government-funded services for Australians.

10.2 The COAG reform agenda — background and rationale

National Competition Policy

The COAG reform agenda can be seen within the context of previous significant reforms on which COAG has embarked. COAG's National Competition Policy (NCP) of 1995 achieved increased competition in Australia through intergovernmental cooperation on micro-economic reforms. The Australian Productivity Commission estimated that the productivity and price changes in key infrastructure sectors in the 1990s, to which the NCP contributed directly, increased Australia's GDP by 2.5 per cent, or \$20 billion (PC 2005, p. xvii).

Human capital

The COAG National Reform Agenda of 2006, which built on the State of Victoria's 'third wave of reform' proposals, sought to address continuing competition challenges, regulatory reform and human capital reform, with the objective of boosting labour force participation and productivity. This was a significant development in cooperation between the Commonwealth and the States. While governments realised that the NCP reforms had been immensely beneficial and needed to continue and deepen, they also realised that a significant wave of human capital reform was required to ensure future prosperity. While competition and

regulatory reforms are intended to make the Australian economy more efficient, human capital reforms lead to a more innovative economy and a more productive workforce. Economic growth and increased income flow from both.

Building on these earlier initiatives, since December 2007 the COAG reform agenda has sought to respond to a number of near and longer-term challenges facing the Australian economy. Globally, the Australian economy is becoming increasingly reliant on the resources sector, and the growth of the Chinese and Indian economies, together with the appreciation of the Australian dollar, will increase international competitive pressures. An ageing Australian population risks reducing overall labour force participation. Productivity increases seen in recent times are unlikely to be sustained unless there is further micro-economic reform.

The COAG reform agenda recognises the importance of continuing the crucial productivity and labour market reforms of the 1990s, but also recognises that human capital reforms are essential for ensuring future prosperity and necessitate better approaches to Commonwealth-State relations. The focus on school education under the COAG reform agenda is one example of COAG's recognition of the importance of human capital reforms. While the States have responsibility for management of the different government schools systems, education is not just a State issue. Education is vital to increasing the productivity of individual workers and the economy as a whole. The Australian Productivity Commission estimated in 2007 that reforms in early childhood, education, skills and workforce development policies could increase productivity by up to 1.2 per cent by 2030 (PC 2006, p. 252). Human capital reform in the area of school education is therefore a national issue and this is reflected in COAG's strategic theme of 'a long-term strategy for economic and social participation'.

Cooperative federalism

Implicit in the COAG reform agenda is the assumption that Australian federalism relies, in many areas, on shared endeavour. The roles and responsibilities of each order of government in Australia should not be oversimplified. In reality, there are a large number of shared areas of policy responsibility. 'Coordinate federalism', as an ideal or pure form of federalism where each order of government does not participate in each other's affairs, will never be possible in Australia. As the former Secretary of the Australian Department of the Prime Minister and Cabinet, Terry Moran, noted:

an enduring and continuing feature of our federation is our shared endeavour in relation to key areas such as health and education. The COAG reform agenda ... [was] explicitly

designed to get more effective outcomes out of shared endeavour, even when governments change and political persuasions differ. (Moran 2010)

In putting the COAG reform agenda proposals to the States through COAG, the Commonwealth made a case for a new model of cooperative federalism and federal financial relations in Australia. The institutional framework that supports the COAG reform agenda, the *Intergovernmental Agreement on Federal Financial Relations* (IGA FFR), reflects the commitment to cooperative federalism explicitly (CRC 2010a, p. 11). It provides that financial relations ‘will be underpinned by a shared commitment to genuinely cooperative working arrangements’ (COAG 2008, p. 5). The IGA FFR envisages governments collaborating ‘on policy development and service delivery’, facilitating the ‘implementation of economic and social reforms’ (COAG 2008, p. 3).

10.3 The COAG reform agenda: key elements

The COAG reform agenda reflects three elements that build on this commitment to cooperative working arrangements and are critical to achieving improved service delivery:

1. funding linked to the achievement of outcomes and outputs (rather than inputs) in areas of policy collaboration
2. devolution of decision making and service design to the frontline wherever possible and effective
3. competitive tensions between the States and Territories (‘competitive federalism’) and competitive tensions between service providers.

Increased transparency and the use of benchmarking to measure performance underpin these three elements, forming a cornerstone of the COAG reform agenda.

Increased transparency and the use of benchmarking to measure performance

As a corollary of the focus on outcomes and outputs and providing the States with increased flexibility, the COAG reform agenda involves increased transparency in funding flows from the Commonwealth to the States, through a streamlined set of publicly available agreements. The COAG reform agenda also provides increased transparency regarding the performance of the States in meeting outcomes, outputs and other targets agreed by COAG. Importantly, it also provides for increased transparency regarding the performance of the Commonwealth in meeting its

commitments in adjacent or related policy areas. For example, the *National Healthcare Agreement* provides for additional transparency relating to primary health care, alongside the much greater transparency relating to the hospital system.

Benchmarking is therefore ingrained in the COAG reform agenda and the supporting institutional framework. This allows the Australian community to know whether the Commonwealth and the States are meeting the agreed outcomes and whether increased Commonwealth funding is achieving improved outcomes.

Benchmarking features of the COAG reform agenda include the establishment of mutually-agreed performance indicators and benchmarks in the National Agreements and National Partnership Agreements, and the assessment of achievement against those benchmarks by the COAG Reform Council (CRC). The CRC is COAG's independent accountability body charged with reporting on performance under the reform agenda. The CRC reports to COAG on the performance of the Commonwealth and the States against the outcomes, outputs and performance benchmarks under the National Agreements, and National Partnership Agreements to the extent they are relevant to the objectives of a National Agreement. The CRC also identifies examples of good practice. As the CRC has noted, the new arrangements are aimed at 'improving performance through fostering and strengthening learning' (McClintock 2010, p. 3).

At its February 2011 meeting, COAG renewed its 'commitment to strong ongoing monitoring and reporting of important national initiatives to ensure that they meet their goals and are delivered in a timely way' (COAG communiqué February 2011, p. 2).

Funding linked with the achievement of outcomes and outputs

The COAG reform agenda emphasises the achievement of outcomes and outputs in areas of policy collaboration, rather than detailed prescriptions by the Commonwealth on how the States will deliver services. Prior to the COAG reform agenda and the accompanying institutional reforms, the States had expressed frustration at the large number of highly prescriptive Commonwealth Specific Purpose Payments to the States. These payments often attached detailed conditions in return for funding, which could hinder States from setting their own priorities in policy and service delivery.

At its November 2008 meeting, COAG stated that the IGA FFR is aimed at 'improving the quality and effectiveness of government services by reducing Commonwealth prescriptions on service delivery by the states, providing them with increased flexibility in the way they deliver services to the Australian people'

(COAG communiqué November 2008). As schedules to the IGA FFR, the six National Agreements between the Commonwealth and the States in key service delivery areas are structured around outcomes and outputs. The National Agreements commit all governments to the achievement of key national objectives, and then provide jurisdictions with the room to tailor policies to meet those objectives while suiting the needs of their own communities, providing more flexibility to spend federal funding within the relevant sector.

The difference between National Agreements and National Partnership Agreements is an important distinction to make, and is perhaps not as well understood as it should be. National Agreements have a purer focus on outcomes and a high degree of autonomy for the States. In contrast, National Partnership Agreements are centred on specific reforms of national priority, projects or service delivery improvements. National Partnership Agreements are intended to be more rigorous in the prescription of specific benchmarks or targets that the States need to achieve to receive Commonwealth funding, and are explicitly intended to set targets that some States do not meet or do not even wish to sign up to. In September 2008, before the IGA FFR had commenced, it was made clear that in return for providing increased funding under National Partnership Agreements (which is additional to base funding under the five National Specific Purpose Payments), the Commonwealth is entitled to seek demonstrated improvements in the delivery of services and clear, measurable outcomes and outputs (Moran 2009).

Devolution of decision making and service design to the frontline

The service delivery frontline is where most Australians interact regularly with their governments. Linking Commonwealth funding with the achievement of outcomes and outputs has moved the Commonwealth away from prescribing in detail how the States should deliver services funded by the Commonwealth. This has, in turn, given the States the opportunity to devolve policy decision-making and service design closer to the service delivery frontline. For example, a key feature of the COAG *National Health Reform Agreement* is the establishment of Local Hospital Networks that will ‘decentralise public hospital management and increase local accountability to drive improvements in performance’ (COAG 2011b, p. 46).

Competitive tensions

Competitive tension between the States is another key element of the COAG reform agenda. The use of reward payments to recognise impressive State performance against pre-determined benchmarks forms a central part of the COAG reform

agenda and fosters a competitive form of federalism. There is approximately \$2.2 billion in reward payments, split among eight National Partnership Agreements, which the Commonwealth committed to provide to the States should they achieve pre-determined performance benchmarks. The CRC assesses independently whether these benchmarks have been achieved. The first reward payments have now been paid in areas such as elective surgery and literacy and numeracy, rewarding States that have delivered what they committed to do.

The CRC's expanded role has helped to inject accountability and performance expectations into the heart of the national debate, which is important for a healthy federation and democracy. This too fosters competitive federalism. For the first time, there is regular and public reporting on whether outcomes, outputs and other targets agreed by all governments are being achieved. CRC reports have attracted considerable media attention. We are already beginning to see a shift in CRC reporting from establishing baselines for measuring performance, to assessing performance over time. It is through this comparative benchmarking of performance over time that the true benchmarking potential of the reforms can be realised — indeed it will only be after 10 years or so of reporting that the CRC's benchmarking will hit its peak impact. Policy learning and service delivery improvements in individual jurisdictions should in turn lead to policy that is more innovative and more responsive to community needs, and therefore to increased levels of healthy competition between the States. Australians will be able to see more clearly which jurisdictions are leading the way in innovative policy development and service delivery improvements, and to what effect.

The benefits of competitive tensions in the Australian federation have been recognised at the State level. The current premier of New South Wales, Barry O'Farrell, has spoken of the need to inject competition into COAG. He has argued that New South Wales will lead 'an agenda that collaboratively defends the value of appropriate national frameworks, but promotes incentives for States to maintain and improve their own competitive advantages' (O'Farrell 2011). The current premier of Victoria, Ted Baillieu, has said that Victoria will seek to pursue a competitive approach to the federation (Dunckley 2010, p. 8). Indeed, the two premiers have gone further and suggested that they will collaborate to drive innovation in areas that are too difficult for all nine jurisdictions to agree (Kenny 2011).

In addition to competition between the States, the COAG reform agenda encourages competitive tensions between service providers, such as individual schools and hospitals. This is achieved by delving below the jurisdictional level and focusing on the organisational dynamics of large service delivery systems, such as education or health systems, managed by the States. Under the COAG reform agenda, the Commonwealth has sought greater transparency from the States in the performance

of individual schools and hospitals. There is now more transparency through the public availability of service level data from the *MySchool* and *MyHospitals* websites. This empowers parents and healthcare consumers with information to make more informed choices, and fosters healthy competition between those service providers.

Remodelling the Commonwealth–State relationship

While the elements outlined above provide a useful focus for understanding the COAG reform agenda, the reform agenda can also be seen as a modern remodelling of the relationship between the Commonwealth and the States. As part of that remodelling, the Commonwealth offered the States a number of things. Introduction of the new framework was coupled with a significant increase in Commonwealth financial support to the States, with COAG agreeing an additional \$7.1 billion over five years in Commonwealth funding associated with the new National Agreements. The Commonwealth offered interventions that were better targeted at specific COAG-agreed reforms (such as through mutually-agreed National Partnership Agreements) as opposed to unilateral Commonwealth interventions in areas of traditional State responsibility. As outlined above, the Commonwealth also offered increased devolution and flexibility in decision-making and service design. Finally, the Commonwealth offered the States a more engaged and collaborative approach to the task of national policy leadership.

In return, the Commonwealth sought from the States much greater levels of transparency, with the States agreeing to be subject to performance reporting by the independent CRC. The Commonwealth also sought more innovation and responsiveness in policy development and service delivery, leading to better uses of Commonwealth funding. Finally, the Commonwealth sought assurances that the States would follow through on COAG-agreed reforms and actually deliver on shared national objectives.

10.4 Progress to date

The COAG reform agenda has made significant progress to date. A comprehensive reform effort is now underway and there have been tangible benefits already. The CRC's 2011 report on the overall progress of the COAG reform agenda (CRC 2011a, p. ix) found that 'governments have made significant progress in realising many of the institutional features of the [IGA FFR]' and that 20 of 26 key reform commitments were 'largely or completely on schedule'.

Institutional successes

Perhaps the most understated benefit of the reforms is the embedding of benchmarking into the reform agenda. The expanded role of the CRC as the ‘key accountability body for the COAG reform agenda’ (CRC 2010a, p. 1) is significant. For the first time, there is regular and public reporting on whether outcomes, outputs and other targets agreed by all governments are being achieved. Vital social policy areas are now receiving the attention they require. There is now regular reporting, on a nationally consistent basis, on outcomes and outputs under the *National Indigenous Reform Agreement* — a significant step forward in the task of establishing higher levels of accountability for Indigenous outcomes. Accountability and performance expectations have been injected into the national policy debate.

In the long term, of course, the real measure of success of the COAG reform agenda will be the extent to which benchmarking and other features of the agenda translate into actual improvements in policy development and service delivery. The shift to a greater focus on outcomes and outputs is intended to ‘focus reform efforts on tangible improvements in the wellbeing of Australians, and to provide governments with the scope to innovate to find the best means of achieving these improvements’ (COAG 2010a, p. 12). CRC reporting on good practice will be important in this area, providing a mechanism for qualitative learning in addition to quantitative performance reporting. As the reform agenda progresses, comparative benchmarking in CRC reports should start flowing back into policy learning and service delivery improvements in individual jurisdictions, prompting policy that is more innovative and more responsive to community needs. However, longitudinal comparison (that is, how a jurisdiction performs over time) is as equally important as horizontal comparison (comparison between jurisdictions). CRC reporting will be important for individual States to see how they are tracking in the long term.

Policy successes

As a result of the COAG reform agenda and unprecedented cooperation between the Commonwealth and the States, there is now better alignment around a number of specific reforms. This alignment includes concrete reform plans in particular policy areas; a clear understanding of shared objectives and outcomes; and better program logic explaining how the Commonwealth and the States will work to achieve those outcomes.

There has also been significant progress in applying micro-economic reform techniques to a number of social policy areas. The National Quality Agenda for Early Childhood Education and Care is an example. Early childhood development is

critical not only for the wellbeing of Australia's children but also for the nation's productivity and workforce participation. Under the new National Quality Agenda, micro-economic reform techniques such as regulatory simplification and establishment of quality benchmarks in early childhood have been utilised in an effort to boost productivity and the wellbeing of Australian children. This has involved replacing nine separate Commonwealth and State systems of licensing, registration, auditing and accreditation, and developing a new single national set of arrangements with new and higher national quality standards. This is a good example of market design, where governments have designed interventions in markets to achieve specific policy outcomes.

Some examples of successes in other social policy areas are outlined below.

School education

The CRC's first report on the overall progress of the COAG reform agenda (CRC 2010a, xii) noted that 'there is a strong focus on reform in the education and skills systems, which should enhance productivity in the long term' (CRC 2010a, pp. xiii-xiv). COAG has endorsed, under the *National Education Agreement*, the development of a national curriculum to replace multiple existing State curricula. A series of National Partnership Agreements, called the Smarter Schools National Partnership Agreements, concentrate on improving teacher quality, better outcomes for low socio-economic status school communities, and improving the essential life skills of literacy and numeracy.

Teacher quality is obviously one of the most important influences on student engagement and achievement. Under the *National Partnership Agreement on Improving Teacher Quality*, \$550 million is being invested to attract the best and the brightest candidates into teaching and to retain quality teachers. The *National Partnership Agreement on Literacy and Numeracy* is providing \$540 million to fund effective evidence-based teaching of literacy and numeracy, and monitoring to identify areas for further support. Under the *National Partnership Agreement on Low Socio-Economic Status School Communities*, \$1.5 billion is being provided to support the educational and wellbeing needs of schools and students in low socio-economic status communities. The *National Partnership Agreement on Literacy and Numeracy* and the *National Partnership Agreement on Improving Teacher Quality* each contain \$350 million in reward funding to reward State performance. Following the first CRC report on the achievement of targets under the *National Partnership Agreement on Literacy and Numeracy*, the Commonwealth announced in June 2011 that it would provide the States with \$138 million in reward funding.

Increased accountability is a key component of COAG's school education reforms and benchmarking has been embedded in the reform package. Under the *National Education Agreement*, jurisdictions committed to greater school transparency, which will allow for better benchmarking of performance. In January 2010 the *MySchool* website was launched, enabling parents and the wider community to compare schools' performance in literacy and numeracy testing against the national average and statistically similar schools. A more advanced version of the website went live in March 2011, which includes summaries of progress made by students in literacy and numeracy since the 2008 national testing, and financial information on schools. This will provide greater insight on the impact of teaching and learning in Australian schools (Gillard 2010) and empower parents with better information.

In November 2011 the CRC released its third annual progress report on the *National Education Agreement*, which includes analysis of performance under National Partnership Agreements that support the objectives of the *National Education Agreement* (CRC 2011b). The report notes that reading and numeracy is improving, although there was mixed progress for Indigenous students.

Elective surgery waiting times

The Commonwealth has made a significant investment in assisting the States to reduce elective surgery waiting times in their public hospitals. All Australians, no matter which State they live in, expect timely public access to elective surgery should the need arise. This is important not only for improved health outcomes but also for patient experience and satisfaction.

As the first step, the Commonwealth entered into a \$600 million Elective Surgery Waiting List Reduction Plan with the States. Up to \$300 million was available under Stage Three of the Plan, which took the form of a National Partnership Agreement between the Commonwealth and the States. The intended outcome of the *National Partnership Agreement on the Elective Surgery Waiting List Reduction Plan* was to reduce the 'number of Australians waiting longer than clinically recommended times for elective surgery by improving efficiency and capacity in public hospitals' (COAG 2009, p. 5). Up to \$252 million in reward funding was available under this National Partnership Agreement. CRC reporting indicated that during the 18 months covered by the Agreement, 54,759 more elective surgery admissions were performed than the 919 389 admissions required under the Agreement (CRC 2011c, p. 10).

The Agreement also contained targets relating to the cost weighted volume of admissions and the management of elective surgery waiting lists, which were assessed by the CRC in its final report. Of the \$252 million in reward funding

available under the National Partnership Agreement, approximately \$144 million was provided by the Commonwealth to the States in recognition of their performance under the Agreement.

More recently, the *National Partnership Agreement on Improving Public Hospital Services* has been developed to implement the elective surgery, emergency department and subacute care elements of the COAG *National Health Reform Agreement*. The National Partnership Agreement invests a further \$800 million in reducing elective surgery waiting times and continues the CRC's role in reporting on whether benchmarks for reward payments have been achieved by the States.

A new performance and accountability framework and improved transparency form a key part of the *National Health Reform Agreement*. This includes benchmarking at the local level, such as reporting on the performance of individual hospitals (and local hospital networks) by the new National Health Performance Authority, and continued benchmarking of jurisdictional performance by the CRC across healthcare services.

The institutional and policy successes outlined above help to illustrate how, on the whole, the Australian federation works well. The former Secretary of the Australian Department of the Prime Minister and Cabinet has argued that the federation 'serves a useful contemporary purpose' and has 'yielded far-reaching policy reforms across many areas, and all levels, of government'. While the promise of the landmark IGA FFR is yet to be fully realised, it 'holds enormous potential for reshaping the delivery of critical services' (Moran 2011).

10.5 Challenges and risks

Despite the successes of the COAG reform agenda, the jury is still out on whether its full potential is being delivered.

Challenges

There is a growing sense at the Commonwealth level that the States have accepted increased levels of Commonwealth funding but are not delivering on their obligations as well as they could be. There is a growing sense at the State level that the Commonwealth is 'reverting to type' and seeking to micromanage the way the States deliver Commonwealth funded services. Balancing the legitimate needs of the Commonwealth and the States is important in making the COAG reform agenda work. In response to Commonwealth requests for performance information, the States may feel the Commonwealth is seeking too much data, too frequently.

Nevertheless, the States need to recognise the legitimate interest of the Commonwealth in ensuring the achievement of agreed reforms or service delivery improvements, including knowing how they will be delivered and the progress that is being made towards agreed outcomes.

Implementation of the reforms and risk management pose challenges for the Commonwealth. One particular challenge, put simply, is who bears responsibility for problems with implementation? As the Commonwealth has greater financial resources than the States, and because Commonwealth Ministers are increasingly expected to engage in public debate on State service delivery, there are frequently expectations that the Commonwealth will intervene in areas in which it has less direct responsibility. This has contributed to the media and the community occasionally holding the Commonwealth to account when implementation falters in State administered programs. While State governments are accountable to their own parliaments and electors for their successes or failures, in some cases the Commonwealth has been under pressure to intervene in areas of program implementation for which it is not directly responsible. An example is implementation difficulties in the Indigenous housing area (Robinson and Franklin 2009).

This context means that Commonwealth Ministers have a high degree of dependence on the performance of the States. In some cases, Ministers may not feel they are given enough information on State progress in implementing COAG reforms to satisfy the cut and thrust of daily political life. This can be a particular challenge in a small number of areas where data limitations inhibit reporting of whether outcomes or other benchmarks are being achieved.

While such concerns are legitimate, it is important that the Commonwealth, as the leading partner in the COAG reform agenda, emphasises the longer-term goals of the COAG reform agenda. While evidence of short-term results is important, especially for tracking progress, it is also important that the media, other stakeholders and even the Commonwealth itself do not lose sight of the longer-term objectives.

This is one aspect of the cultural change that is required if governments are to stay the course of the reforms (CRC 2010a). The IGA FFR requires a significant shift in the Commonwealth bureaucracy's instinct to prescribe in detail how the States will deliver services funded by the Commonwealth. The Commonwealth needs to accept ways of managing risk other than through input controls, such as better use of reward payments and utilising the CRC's comparative reporting of State performance. Such cultural change is essential not only in Commonwealth central

agencies but especially in line departments, which are at the coalface of interaction with State line agencies on program delivery.

Risks

There are clear risks to the COAG reform agenda. These risks are compounded because the jury is still out on whether the full potential of the COAG reform agenda is being delivered. In a period of fiscal consolidation, it will be difficult for the Commonwealth to preserve current funding levels for a number of initiatives when the agreements governing these initiatives expire. Some policy debates, such as health reform, have explicitly moved to a related but separate institutional framework. For example, under the COAG *National Health Reform Agreement*, funding for hospitals will be contributed into a single national pool, to be operated by an independent Administrator and supported by a new National Health Funding Body. Commonwealth funding contributions to public hospital services will be provided on the basis of actual activity levels, measured and reported regularly (COAG 2011b). This is clearly a significant contrast to the original model for healthcare collaboration envisaged under the IGA FFR and the *National Healthcare Agreement*.

Unless some of these risks dissipate, the Commonwealth will find it difficult to continue to offer the States the existing levels of flexibility in policy and service delivery design that the COAG reform agenda has provided. Under pressure for faster and better delivery, there is a range of alternative proposals for reform in key policy areas that the Commonwealth could adopt, which would give the Commonwealth much greater policy control. These would offer much less flexibility to the States but might be seen by Commonwealth Ministers to deliver more to the nation.

Despite the need for shared endeavour and cooperation, the historic trend in the Australian federation has been towards centralism, mainly due to the high degree of vertical fiscal imbalance. Fenna (2007, p. 298) notes there is general agreement that Australia is the most centralised of the established federations, and that the underlying trend 'is toward centralisation rather than decentralisation'. While the COAG reform agenda has helped to institutionalise a new cooperative form of federalism, at least one commentator has argued (Anderson 2010, p. 17) that this has not resulted in any significant change or slowing 'in the development of the Australian federation towards a model of a strong central government setting priorities and determining policies, which then funds the states to implement the programs required by those policies'.

The States play a critical role in the federation. However, it is indisputable that aspects of that role have eroded over time. The COAG reform agenda offers the States a modern 21st century opportunity to reinvent and reinvest in their strengths. The Commonwealth has put significant effort and resources into giving the States increased opportunities and flexibility to improve outcomes for their communities. If the States fail to deliver improved services, it is unlikely the trend towards centralisation described above can be arrested.

10.6 Conclusion and next steps

For the federation to remain relevant, the COAG reform agenda needs to succeed. And for believers in Australian federalism, now is the time for delivery. Barring major reallocations of roles and responsibilities within the federation, achieving nationally significant reforms will continue to require collaboration between the Commonwealth and the States. Funding linked to outcomes and outputs, greater devolution, competitive federalism, and increased transparency and benchmarking will all remain critical to improving government-funded services for Australians.

To enable the COAG reform agenda to succeed, the Commonwealth needs to get much better at a collaborative model of national policy leadership, put greater pressure on the States to live up to their obligations, and attempt to avoid the instinctive desire to prescribe in detail how the States should deliver services that are Commonwealth-funded. The States need to follow through on COAG-agreed reforms and ensure they actually deliver shared national objectives. All jurisdictions need to recognise that change will take some time and that policy consistency over time is a virtue in the federation.

All governments need to work together to produce better data. This includes more investment in data collection and manipulation; more focus on the operations and effectiveness of key data agencies; and more exploration of the links between the performance reporting framework underpinning the COAG reform agenda and broader frameworks both in Australia and internationally, such as the OECD ‘measuring the progress of societies’ agenda. Effective public accountability is dependent on jurisdictions providing robust data comparable between jurisdictions. While often more difficult to collect, data showing whether outcomes are being achieved (in addition to outputs) are also critical. COAG is already acting to address data challenges, including by reviewing the performance frameworks in the National Agreements with the objective of ensuring that ‘progress is measured and that all jurisdictions are clearly accountable to the public and COAG for their efforts’ (COAG communiqué February 2011, p. 2).

Most importantly, governments and their officials need to work together on policy innovation mechanisms to embed benchmarking and competitive federalism, and turn this into innovative federalism. The ‘missing glue’ of the COAG reform agenda is the connection between data and policy impact in areas such as policy innovation, best practice, competitive or ‘laboratory’ federalism and policy markets. These are underdeveloped in Australia and warrant further investment. Better support to enable policy innovation is needed. Public servants and politicians in the Commonwealth and the States have responded to the challenges of the COAG reform agenda through innovation, but there is room for greater support and facilitation of that policy innovation. This includes providing greater opportunities for collaboration and fostering creativity. At the Commonwealth level, these matters are being addressed as part of the Government’s response to the report *Ahead of the Game: a blueprint for reform of Australian Government administration* (Advisory Group on Reform of Australian Government Administration 2010).

Greater support for policy innovation will enable the potential of the COAG reform agenda to be realised more fully. This will lead to a more innovative federation and, ultimately, better services for Australians.

References

- Advisory Group on Reform of Australian Government Administration 2010, *Ahead of the Game: a blueprint for reform of Australian government administration*, Department of the Prime Minister and Cabinet, Canberra.
- Anderson, G. 2010, ‘Wither the Federation? Federalism under Rudd’, *Public Policy*, vol. 5, no.1, pp. 1–22.
- COAG Reform Council 2010a, *COAG Reform Agenda: Report on Progress 2010*, Sydney.
- 2010b, *National Partnership Agreement on the Elective Surgery Waiting List Reduction Plan: Period 1 and Period 2 Assessment Reports*. Sydney.
- 2011a, *COAG Reform Agenda: Report on Progress 2011*, Sydney.
- 2011b, *Education 2010: comparing performance across Australia (National Education Agreement Performance Report)*, Sydney.
- 2011c, *National Partnership Agreement on the Elective Surgery Waiting List Reduction Plan: period 3 assessment report*, Sydney.
- COAG 2008, *Intergovernmental Agreement on Federal Financial Relations*, Council of Australian Governments, Canberra.

-
- 2009, *National Partnership Agreement on the Elective Surgery Waiting List Reduction Plan*.
- 2010, *National Health and Hospital Network Agreement – National Partnership Agreement on Improving Public Hospital Services*.
- 2007–2010, Meeting Communiqués.
- 2011a, *Heads of Agreement on National Health Reform*.
- 2011b, *National Health Reform Agreement*.
- Craven, G. 2009, ‘The Future of Federalism’, *Australasian Parliamentary Review*, vol. 24, no. 2, pp. 23–27.
- Department of the Prime Minister and Cabinet 2010, *Meeting the Challenge: COAG’s reform agenda*. Unpublished.
- Dunckley, M. 2010, ‘It’s Victoria against the rest: Baillieu’, *The Australian Financial Review*, 3 December, p. 8.
- 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.
- Gillard, J. 2010, *Media Release – Providing unprecedented school performance data* (17 November 2010), www.pm.gov.au/print/node/7014 (accessed 20 November 2010).
- 2011, ‘Four themes for a better COAG’, *Australian Financial Review*, 8 February.
- Kenny, C. 2011, ‘Liberal Premiers Barry O’Farrell and Ted Baillieu flex their collective muscle’, *The Australian*, 15 December.
- McClintock, P. 2010, *The COAG Reform Agenda: how are governments performing so far?* Speech, 21 September. http://www.coagreformcouncil.gov.au/media/speeches/speech_20100921_pdf (accessed 20 November 2010).
- Moran, T. 2009, *Splicing the Perspectives of the Commonwealth and States into a Workable Federation*, in J. Wanna (ed.), *Critical Reflections on Australian Public Policy: selected essays*, ANU e-press, Canberra.
- 2010, *The Future of the Australian Public Service: challenges and opportunities*. Neil Walker Memorial Lecture, 13 October, Department of the Prime Minister and Cabinet, Canberra. www.dpmmc.gov.au/media/speech_2010_10_14.cfm (accessed 20 November 2010).
- 2011, *The Challenges of Federalism*, Address to the Eidos Institute, 8 June. Department of the Prime Minister and Cabinet, Canberra. www.dpmmc.gov.au/media/speech_2011_06_08.cfm (accessed 9 June 2011).

O'Farrell, B. 2011, Speech to the National Press Club, 9 February.

PC (Productivity Commission) 2005, *Review of the National Competition Policy Reforms*, Report no. 33, Melbourne.

— 2006, *Potential Benefits of the National Reform Agenda*, Report to the Council of Australian Governments, Melbourne.

Robinson, N. and Franklin, M. 2009, 'Stark call to end housing "gravy train"', *The Australian*, 6 August.

Silver, H. 2010, 'Getting the Best out of Federalism – the Role of the Productivity Commission and the Limits of National Approaches', *Australian Journal of Public Administration*, vol. 69, no. 3, pp. 326–332.

11 Benchmarking and accountability: the role of the COAG Reform Council

Mary Ann O’Loughlin¹
COAG Reform Council

Reforms in 2008 made benchmarking a central element of federal financial relations in Australia. As implemented by the COAG Reform Council, this involves assessing the performance of the Commonwealth and State governments in achieving outcomes and benchmarks in areas of nationally significant reforms. Making such a system work requires reliable data; agreement on meaningful targets; recognition that the States carry primary responsibility for program design and implementation; and ongoing commitment to effective intergovernmental collaboration.

11.1 The Intergovernmental Agreement on Federal Financial Relations

To understand the benchmarking roles of the COAG Reform Council, it is essential to understand the context in which they are defined.²

The Council’s roles and responsibilities are set out in the Intergovernmental Agreement on Federal Financial Relations, an historic agreement that provides an overarching framework for the Commonwealth government’s financial relations with the States and Territories (COAG 2008a, p. 2). The agreement was signed by the Prime Minister, Premiers and Chief Ministers at a meeting of the Council of Australian Governments (COAG), and took effect on 1 January 2009.

According to COAG (2008a, p. 2), the Intergovernmental Agreement:

¹ Mary Ann O’Loughlin is Executive Councillor and Head of Secretariat of the COAG Reform Council.

² The COAG Reform Council was established in 2006 but its roles and responsibilities were significantly expanded in 2008. This article discusses the roles of the Council under its expanded mandate.

represents the most significant reform of Australia's federal financial relations in decades. It is aimed at improving the quality and effectiveness of government services by reducing Commonwealth prescriptions on service delivery by the States, providing them with increased flexibility in the way they deliver services to the Australian people. In addition, it provides a clearer specification of roles and responsibilities of each level of government and an improved focus on accountability for better outcomes and better service delivery.

The Intergovernmental Agreement is an agreed approach — or set of approaches — for addressing two key features of Australia's federal system (see Banks, Fenna and McDonald, this volume). First is the problem of vertical fiscal imbalance (VFI) — the States have large expenditure responsibilities relative to their revenue raising capacities, and hence must rely on financial transfers from the Commonwealth. The second is that the Commonwealth and State governments have overlapping roles and responsibilities for service delivery, including for healthcare, disability services, housing, and education.

There are three main elements of the new financial arrangements: National Specific Purpose Payments supported by new National Agreements; National Partnership Payments associated with National Partnership Agreements; and a performance and assessment framework to support public reporting and accountability.

National Specific Purpose Payments and National Agreements

Under the new framework for federal financial relations, the previous more than 90 different payments from the Commonwealth to the States for specific purposes — many containing prescriptive conditions on how the funding should be spent — have been combined into five new National Specific Purpose Payments (Commonwealth of Australia 2009, p. 24). National Specific Purpose Payments are ongoing financial contributions from the Commonwealth to the States to be spent in the key service delivery sectors of schools, skills and workforce development, health care, affordable housing, and disability services. The States are required to spend each National Specific Purpose Payment in the service sector relevant to the payment, but they have full budget flexibility to allocate funds within that sector as they see fit to achieve the agreed objectives for that sector (COAG 2008b, p. D-2).

National Specific Purpose Payments are associated with National Agreements between the Commonwealth and State governments. National Agreements establish the policy objectives in the service sectors of education, skills and workforce development, health care, affordable housing, and disability services. There is also a National Agreement on Indigenous Reform which does not have an associated

Specific Purpose Payment, although it links to other National Agreements and National Partnerships which have associated funding.

National Agreements set out the objectives, outcomes, outputs and performance indicators for each sector, which are agreed between all jurisdictions. The agreements also aim to clarify the roles and responsibilities of the Commonwealth and States in the delivery of services and the achievement of outcomes. They do not include financial or other input controls imposed on service delivery by the States, and there is no provision for National Specific Purpose Payments to be withheld in the case of a jurisdiction not meeting a performance benchmark specified in a National Agreement.

National Partnership Agreements and payments

National Partnership Agreements outline agreed policy objectives in areas of nationally significant reform or for service delivery improvements, and define the outputs and performance benchmarks. They cover a wide range of service sectors and reform areas, from health and education through to regulation and competition reform.³

National Partnerships differ from National Agreements in that generally they are time-limited and the associated National Partnership payments for the States are linked with specific reform activities or projects. The Commonwealth provides National Partnership payments for three purposes: to support the delivery of specified projects, to facilitate reforms, or to reward those jurisdictions that deliver on national reforms (Commonwealth of Australia 2009, p. 26).

Performance and assessment framework

The third main element of the new federal financial relations arrangements is a performance and assessment framework to support public reporting and accountability. Under the *Intergovernmental Agreement*, the Commonwealth and States have committed to greater accountability through simpler, standardised and more transparent performance reporting, and ‘a rigorous focus on the achievement of outcomes — that is, mutual agreement on what objectives, outcomes and outputs improve the well-being of Australians’ (COAG 2008b, pp. 5–6). The *Intergovernmental Agreement* gives the COAG Reform Council significant responsibilities for assessment and reporting of the performance of governments under National Agreements and National Partnerships.

³ A list of current National Partnerships is at <http://www.federalfinancialrelations.gov.au/>.

11.2 Role of the COAG Reform Council

The COAG Reform Council assists COAG to drive its national reform agenda by strengthening accountability for the achievement of results through independent and evidence-based monitoring, assessment and reporting on the performance of governments. The Council is funded by all governments but is independent of individual governments and reports directly to COAG. The Commonwealth government appoints the Chairman of the Council, the States appoint the Deputy Chairman, and the governments jointly appoint the other four members. At least one member must have regional and remote experience. There is also an Executive Councillor and head of the secretariat.⁴

As set out in the *Intergovernmental Agreement*, the COAG Reform Council has two main benchmarking roles related to National Agreements and National Partnerships (COAG 2008b, p. A-4). The Council independently assesses and publicly reports on:

- the performance of the Commonwealth and States in achieving the outcomes and benchmarks specified in National Agreements
- whether performance benchmarks in nationally significant reforms have been achieved before the Commonwealth government makes reward payments to the States.

These roles are described below, followed by a discussion of some key challenges faced by the COAG Reform Council in its early years of reporting.

11.3 Benchmarking under the National Agreements

For each of the six National Agreements, the COAG Reform Council provides annual reports to COAG based on a comparative analysis of the performance of governments against indicators that have been agreed by the governments. The first year reports establish baseline data against which progress in reform and improvements in service delivery can be measured (COAG 2008b, p. C-3). The reports are made public.

The performance information for each National Agreement is received from the Steering Committee for the Review of Government Service Provision, a cross-jurisdictional body (see Banks and McDonald this volume). The Steering Committee collates the data from relevant sources and provides the data to the

⁴ An overview of the role of the COAG Reform Council is at <http://www.coag.gov.au/crc/index.cfm>.

Council. The Council's report on the National Agreement is due to COAG within three months of receiving the data. The Council must formally consult with the jurisdictions on the report during this three month period.

Example of the National Education Agreement

As an illustrative example, we may take the *National Education Agreement* (which is discussed further by Glover and Dawkins in this volume). The *National Education Agreement* is supported by a National Schools Specific Purpose Payment of about \$11.4 billion (in 2011-12).

Structure

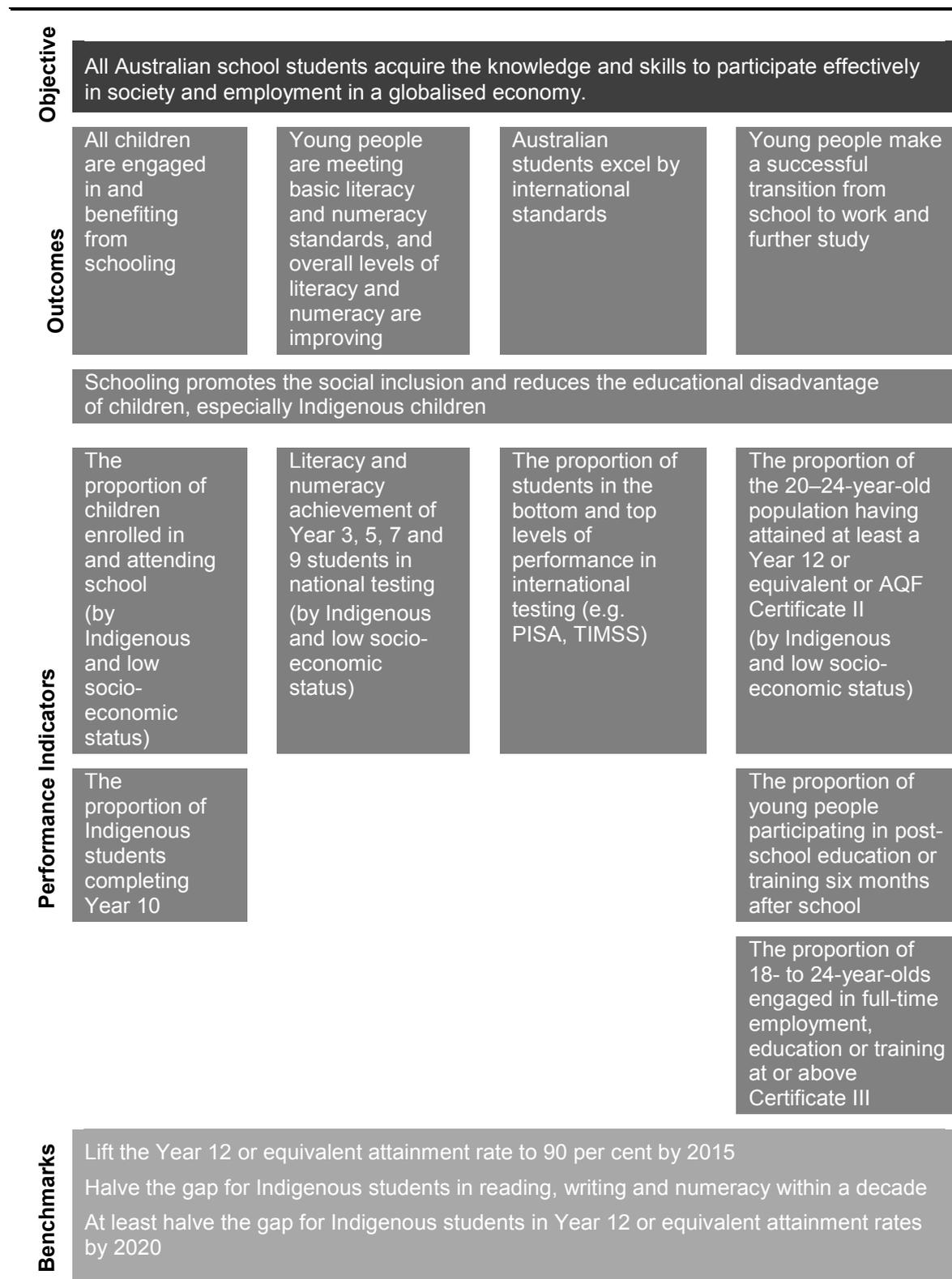
Figure 11.1 summarises the structure of the *National Education Agreement*. The six National Agreements — in education, skills and workforce development, health care, affordable housing, disability services and Indigenous reform — have a similar structure. All National Agreements begin with the objective(s) of the agreement — the overall aim. The objective of the *National Education Agreement* is:

that all Australian school students acquire the knowledge and skills to participate effectively in society and employment in a globalised economy. (COAG 2008c, p. 1)

Each agreement also has a set of outcomes agreed by governments. As shown in figure 11.1, the National Education Agreement has five outcomes. For each outcome there is a set of performance indicators which measure progress towards the outcomes. For example, for the outcome of 'young people meet basic literacy and numeracy standards and that levels of achievement are improving', the performance indicator is literacy and numeracy achievement of Year 3, 5, 7 and 9 students in annual national testing under the National Assessment Program — Literacy and Numeracy (NAPLAN).

Most National Agreements also identify performance benchmarks to be achieved. The *National Education Agreement* has three (listed in figure 11.1).

Figure 11.1 Structure of the National Education Agreement



Source: COAG (2008c).

Baseline reporting

Each year, the COAG Reform Council publicly reports the performance information and undertakes a comparative analysis of the performance of jurisdictions towards the outcomes, as measured by the performance indicators and benchmarks (COAG 2008b, p. C-2). The first year reports establish the baselines for performance. The comparative analysis compares the performance of jurisdictions against each other and also against their own year-on-year performance, reflecting the importance of achieving continuous improvement against the outcomes, outputs and performance indicators.

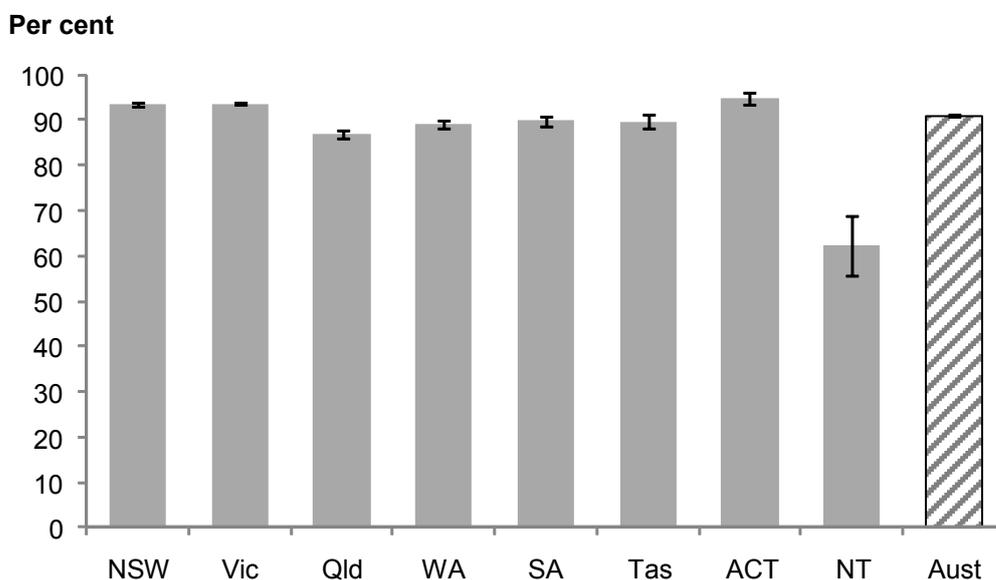
To take an example, figure 11.2 presents the baseline data (2008) for the States against one of the performance indicators of the National Education Agreement: ‘young people meet basic literacy and numeracy standards’. The agreed measure of the indicator is the proportion of students achieving at or above the national minimum standard. Achievement of the minimum standard indicates that the student has demonstrated the basic elements of literacy and numeracy for the year level.

The data against this indicator are shown for Year 5 Reading. Year 5 Reading is a good indicator of performance, as reading is a foundation skill for writing and numeracy and by Year 5 the impact of jurisdictional differences in school starting age on the acquisition of skills should be diminishing.

Nationally, a high proportion of all students — 91 per cent — achieved at or above the national minimum standard in assessments of reading at Year 5 in 2008 (COAG Reform Council 2009a, pp. 61–2). Comparing the performances of the States:

- three jurisdictions achieved higher levels than the national average — New South Wales, Victoria and the Australian Capital Territory — all with results of 94 per cent or above of students meeting the national minimum standard in Year 5 Reading
- four jurisdictions clustered below the national average, with the proportion of students meeting the national minimum standard ranging from 90 per cent in South Australia and Tasmania, to 89 per cent in Western Australia, and to 87 per cent in Queensland
- the Northern Territory differed significantly from other States, with 63 per cent of students meeting the national minimum standard for Year 5 Reading.

Figure 11.2 Proportion of Year 5 students achieving at or above the national minimum standard for reading, by State, 2008 ^{a,b,c}



^a The achievement percentages shown in this graph include 95 per cent confidence intervals indicated by error bars. ^b Exempt students were not assessed and are deemed not to have met the national minimum standard. ^c Absent and withdrawn students did not sit the test and are not included in these data.

Source: MCEETYA (2008, p. 56).

Contextual differences between jurisdictions

To help understand performance, the Council is also required to highlight contextual differences between the jurisdictions which are relevant to interpreting the data, such as relevant demographic characteristics (COAG 2008b, p. C-2). For example, the Northern Territory has a high proportion of Indigenous students (41 per cent). Indigenous children are the most educationally disadvantaged group in Australia and this disadvantage is reflected in most measures of educational outcomes.

The Council’s approach to the task of highlighting contextual differences reflects its general approach to the assessment of governments’ performance (COAG Reform Council 2009a, pp. 6–7). In particular, the Council’s approach is dynamic, emphasising changes in performance from year to year.

Given this approach, the contextual differences that are highlighted in assessing performance under the National Agreements are high level and small in number. They are focused on differences that help interpret the data by giving the broad context, in particular student characteristics such as Indigenous and socioeconomic status. This is particularly relevant for first-year reports, as they present the baseline

data for the comparative assessment of performance. However, these contextual factors are likely to be less relevant to understanding changes in performance. They are, hence, less likely to be relevant in subsequent years' reports as the focus shifts to assessing the performance of the jurisdictions over the years compared with their baseline data (COAG Reform Council 2009a, p. 26).

Analysing change over time

With the baseline data published, the second and subsequent year reports shift to assessing governments' progress against agreed objectives, outcomes and outputs. The shift to assessing progress means a focus on assessing change over time.

'Change over time' can be described as progress, improvement, decline or failure to progress, depending on the direction of change and other considerations. Within the Council's comparative analysis framework, change over time is a dynamic construct as it involves analysing change within and across jurisdictions, and for key sub-populations where possible.

Table 11.1 gives an example of change over time analysis. For each jurisdiction, it shows the change in the proportion of Indigenous students at or above the national minimum standard in Reading in Years 3, 5, 7 and 9 between 2008 and 2011. In summary:

- Nationally, in Years 3 and 7, the proportion of Indigenous students at or above the national minimum standard in Reading was significantly higher in 2011 than in 2008. There was no significant change in Years 5 or 9.
- The most notable feature of the data is the lack of significant progress in the proportion of Indigenous students at or above the national minimum standard in Reading in most States over the four years.
- Comparing the States, only Queensland and Western Australia had significant increases in achievement in Reading, with both significantly improving in Years 3 and 7.
- Both New South Wales and Tasmania had significant decreases in achievement in Reading in Year 9.

Table 11.1 Summary of significant changes in the proportion of Indigenous students at or above the national minimum standard in Reading, 2008-2011

	Year 3	Year 5	Year 7	Year 9
New South Wales	—	—	—	▼
Victoria	—	—	—	—
Queensland	▲	—	▲	—
Western Australia	▲	—	▲	—
South Australia	—	—	—	—
Tasmania	—	—	—	▼
Australian Capital Territory	—	—	—	—
Northern Territory	—	—	—	—
Australia	▲	—	▲	—

▲ statistically significant increase in achievement. ▼ statistically significant decrease in achievement. — no statistically significant change.

Source: COAG Reform Council (2012, p. 48).

11.4 Benchmarking under Reward National Partnerships

The COAG Reform Council's main benchmarking role for National Partnerships is to independently assess whether performance benchmarks or milestones have been achieved before the Commonwealth makes reward payments to the States.

As of March 2012, there were seven National Partnerships with reward payments agreed by all jurisdictions (table 11.2). As shown, reward National Partnerships differ in terms of performance measures, reward funding and reporting timeframes. The predetermined performance benchmarks are variously included in National Partnership Agreements, or contained in greater detail in implementation plans. Implementation plans are developed by jurisdictions and subject to approval by the relevant Commonwealth Minister.

In terms of processes, the Council makes a distinction between the six reward National Partnerships in health and education, and the National Partnership for a Seamless National Economy, as explained below.

Table 11.2 Reward National Partnerships ^a

<i>National Partnership</i>	<i>Benchmarks</i>	<i>Reward funding</i>	<i>Reporting dates</i>
Literacy and Numeracy	Literacy and numeracy benchmarks for students most in need of support, especially Indigenous students	\$350m	March 2011 April 2012
Youth Attainment and Transitions	Benchmarks for participation in education and attainment of Year 12 or equivalent	\$100m	August 2012* April 2013
Improving Teacher Quality	Reform benchmarks and milestones	\$350m	April 2012 April 2013*
Improving Public Hospital Services	Benchmarks for elective surgery and access to emergency departments.	\$400m	2012-13 to 2015-16
Essential Vaccines	Benchmarks for coverage and wastage and leakage	\$24m	2011 and ongoing
Preventive Health	Benchmarks for healthy bodyweight, fruit and vegetable consumption, physical activity and smoking	\$308m	November 2013* May 2015*
Seamless National Economy	Milestones for competition and regulatory reform	\$450m	December 2008 to 2013

^a Reward National Partnerships as of March 2012. * Date to be confirmed.

Reward National Partnerships in health and education

While each of the six National Partnerships in health and education with reward payments specifies a role for the Council, the language and specific requirements differ. In consultation with the jurisdictions, the Council developed a common process and transparent set of principles for assessment.

A key step in the process is the development of a matrix of performance information for each National Partnership with reward funding prior to reporting on the National Partnership. The matrix of performance information establishes the specific framework for assessment, clearly setting out the basis for assessment, consultation arrangements, and reporting timeframes.

The matrix is circulated to jurisdictions for one month consultation prior to the commencement of each assessment period.

National Partnership for a Seamless National Economy

The National Partnership Agreement to Deliver a Seamless National Economy has separate assessment and reporting arrangements.

Under the National Partnership, the Commonwealth and the States have agreed to work together:

to deliver more consistent regulation across jurisdictions and address unnecessary or poorly designed regulation, to reduce excessive compliance costs on business, restrictions on competition and distortions in the allocation of resources in the economy. (COAG 2008b, p. 3)

There are 36 agreed streams of regulation and competition reform. The National Partnership is underpinned by an implementation plan that ‘articulates the policy outcomes sought in each reform area and, where possible, also identifies key milestones for jurisdictions in progressing each reform’ (COAG 2008e, p. 5).

The COAG Reform Council reports annually to COAG, providing an independent assessment of whether the milestones in the National Partnership have been achieved. The Council’s first report on the performance of governments under the National Partnership was presented to COAG in December 2009 (COAG Reform Council 2009b).

The National Partnership provides for reward payments of up to \$450 million over 2011-12 and 2012-13 for delivery of the 27 deregulation priorities. The Council advises on achievement of key milestones for deregulation priorities before reward payments are made. States are eligible for full payment even if one reform is not met, as long as it is not one of the 13 priorities.

The Council’s assessment of performance is evidence-based and draws on a range of inputs. These include progress reports from the jurisdictions to the Council; additional information requested by the Council to assist the assessment process; and the Council’s independent research on legislative and regulatory activities of governments.

The Council uses a green–amber–red ‘traffic light’ representation of progress against individual milestones. The Council also undertakes an assessment of risks to the achievement of future milestones and explores broader risks to the achievement of the intended output of the reform stream.

11.5 Key challenges

In the early years of reporting, the Council faced a number of challenges undertaking its benchmarking roles. The following sections highlight some of the key challenges in reporting against National Agreements and reward National Partnerships.

Outcomes-based accountability

Under the *Intergovernmental Agreement*, the Commonwealth moved away from prescriptive tied grants. In their place it put a system of block grants linked to National Agreements and independent assessment of performance by the COAG Reform Council. The *Intergovernmental Agreement* is clear that it is a shift away from accountability based on inputs and outputs to outcomes-based accountability.

The question of whether these new financial arrangements are (or will with time) be successful turns critically on whether outcomes have (or will) improve. In turn, how well the new arrangements succeed in improving outcomes depends critically on the effectiveness of the incentives to encourage governments to take action.

While there are Commonwealth payments to the States associated with National Agreements, there is no provision for payments to be withheld on the basis of the State's performance against the performance indicators or benchmarks. Hence, the incentives for improved performance flow from the potential gain or loss to a State's reputation from the public reporting of its performance, particularly in comparison to other jurisdictions and over time.

For comparative analysis to be an effective incentive for improved performance it is essential that there a strong performance reporting framework — that is, the agreed objectives, outcomes and performance indicators and the associated information and data against which the Council makes its assessments. The Council must be able to assess and compare the progress of each jurisdiction over time in the areas covered by the National Agreements. It must, therefore, have access to adequate and reliable information and data to inform its assessments.

Unusual for an agreement on federal financial relations, the *Intergovernmental Agreement* recognises the importance of performance reporting. To quote from the Agreement:

the success of the new framework for federal financial relations depends crucially on the development of robust performance indicators and benchmarks. (COAG 2008b, p. C-5)

In the early years of reporting under the National Agreements, the Council faced significant challenges associated with the performance reporting frameworks for the agreements. In its reports to COAG, the Council has urged major improvements to the performance reporting framework, particularly in two areas.⁵

⁵ See: http://www.coagreformcouncil.gov.au/media/speeches/speech_170910_natstats.pdf.

First, the Council has called for improvements to the conceptual adequacy of indicators. Across the National Agreements, there are problems with some indicators not being closely connected to the objectives and outcomes of the agreements. The Council reinforces the view of COAG in the *Intergovernmental Agreement on Federal Financial Relations* that:

The purpose of the performance indicators is to inform the general public about government performance in meeting progress towards identified outcomes. (COAG 2008b, p. C-2)

To do this requires performance reporting for National Agreements to be based on a strong conceptual framework that links the performance indicators with the outcomes. Performance indicators should provide a clear picture of achievement.

The second area where the Council is urging significant improvements is in the availability of adequate data for reporting progress against performance indicators and benchmarks. There are a number of data limitations that have reduced the effectiveness of performance reporting. These include poor quality and unreliable data; data that are not comparable over time or between jurisdictions; and data that cannot be sufficiently disaggregated by Indigenous or socio-economic status. All National Agreements have examples of performance indicators which have no data or have inadequate data to report progress; for all National Agreements there are also significant problems with data to report progress over time.

Partly in response to the Council's recommendations on improving the conceptual adequacy of indicators and the availability of adequate data, COAG agreed to review the performance reporting frameworks of the National Agreements. The reviews will cover the conceptual adequacy of the performance reporting frameworks, the appropriateness of performance indicators and the availability of adequate data. The reviews are scheduled to be completed by the end of June 2012.

Setting clear and ambitious benchmarks

The *Intergovernmental Agreement* is very clear that the Commonwealth and States should agree on ambitious benchmarks and milestones under National Partnerships. It states:

National Partnerships should set out clear mutually agreed and ambitious performance benchmarks... that encourage achievement of ambitious reform targets and continuous improvement in service delivery, and provide better outcomes than would otherwise be expected. (COAG 2008b, p. E-3)

The Council has, however, been critical of the level of clarity and ambition reflected in many benchmarks and milestones under National Partnerships.

For instance, in reporting on the National Partnership Agreement on a Seamless National Economy, the Council argued for a more coherent set of milestones and of more rigorous specification of milestones and deadlines for assessing progress in a number of reform areas (COAG Reform Council 2009b; 2010b).

In many reward National Partnerships in health and education there is the added challenge that, while the National Partnerships provide an overarching multilateral framework, each State is accorded flexibility to implement reform strategies most appropriate to their government's policy settings and circumstances. States develop their own detailed implementation plans — agreed bilaterally with the Commonwealth — outlining the reforms they intend to introduce and their benchmarks and measures for assessment (COAG Reform Council 2011b, p. 86).

This recognition of State differences in policy settings and circumstances is important in a federal system. However, for each National Partnership, this can result in a high level of variation between States in reward frameworks and varying levels of ambition across the States in determining benchmarks for improvements.

For the Council, this presents difficulties for public accountability. While it is the Council's role to assess achievement of the agreed performance benchmarks, the Council is not mandated to assess the level of ambition or degree of difficulty associated with achieving the benchmarks (COAG Reform Council 2011b, p. 89). Thus, the fact that a State is assessed as meeting all its agreed benchmarks does not necessarily mean that it has achieved more than another jurisdiction which failed to meet its benchmarks. It is possible that the second State set more ambitious benchmarks and succeeded in improving outcomes more than the first State, even though it missed meeting its benchmarks.

The Council has argued that, while flexibility is important, it is essential that there is a degree of comparability of frameworks and ambition if benchmarking is to be an effective tool of public accountability. It has recommended for all future National Partnerships COAG considers that performance benchmarks for each State should be independently assessed for ambition and the assessment made publicly available (COAG Reform Council 2012, p. 68).

Catalyst data

Even with a robust performance reporting framework and data, a comparative analysis does not explain why there are differences between the jurisdictions or why performance has improved or declined over time. As Fenna (this volume) notes, rather than providing an explanatory analysis, the comparative analysis is better thought of as providing catalyst data (Ekholm 2004, p. 1). The comparative analysis

of performance — highlighting differences among the jurisdictions or over time — leads one to search for reasons to explain the differences. It does not provide the ‘right’ answers or answer questions about why programs work or fail. But performance information can signal that something is wrong — or right — and prompt debate. It can encourage governments to consider what to do to improve (Commonwealth of Australia, Department of Finance and Deregulation 2012, p. 50).

It is too early to assess the effectiveness of the Council’s comparative assessments under the National Agreements in encouraging governments to take action to improve their performance. But catalyst data can be a powerful incentive. The state of Queensland provides an example of this, as pointed out by an editorial in *The Australian* newspaper on Queensland’s results in national testing in literacy and numeracy:

Following the woeful performance of Queensland primary school children in national testing last year, the Bligh Government turned to an expert for help. The state’s children need it, after being ranked second-last in the nation. Only the Northern Territory where absenteeism and social disadvantage are more prevalent fared worse. (*The Australian* 4 May 2009, editorial)

In response to Queensland’s performance in national testing in 2008, Premier Bligh took remedial action, seeking an independent review by Professor Geoff Masters from the Australian Council for Educational Research of the literacy and numeracy standards in Queensland primary schools and advice on how to improve students’ skills (Masters 2009). Since then, the Council has noted that Queensland has more consistently improved performance in national testing than other jurisdictions (COAG Reform Council 2010a, p. 19).

Highlighting good practice

While the Council is not tasked with explaining the differences between the jurisdictions, it does have a role under the Intergovernmental Agreement to highlight examples of good practice and performance so that, over time, innovative reforms or methods of service delivery may be adopted by other jurisdictions (COAG 2008b, p. C-3). In 2010-11, the Council completed research projects in each of the six National Agreement reform areas. The projects looked at variations in relative performance across the jurisdictions for selected indicators to help identify possible areas for good practice analysis.⁶

⁶ The reports are available at www.coagreformcouncil.gov.au/excellence/good_practice.cfm.

In 2012, the Council commenced a series of conferences to bring together the jurisdictions, researchers and interest groups to discuss examples of good practice within Australia and internationally in areas of nationally significant reform. The Council selected the area of transitions from school as the focus of the first of these forums, investigating ‘what works’ in keeping young people in education, employment and/or training.

11.6 Lessons learned

At the time of writing this article, the benchmarking arrangements under the Intergovernmental Agreement are still fairly new, having been in place for little more than three years. From this early period, three lessons stand out.

The first is that for the benchmarking arrangements to be effective they must be based on robust performance reporting frameworks, which are conceptually sound and supported by quality, comparable and timely performance information. Progress should also be assessed against clear milestones and outcomes and ambitious benchmarks. The aim is to encourage — even pressure — governments to take action in response to performance feedback.

Second, it is important to get the right balance between flexibility for States to determine their own priorities and accountability to ensure that the objectives of funding agreements are being achieved. The Intergovernmental Agreement recognises the strengths of federalism, and in particular the primacy of the States in service delivery, by focusing on outcomes and flexibility rather than prescribing a one-size-fits-all approach. But it also seeks to address the challenges of federalism through clearly defined roles and responsibilities and clear accountabilities. The question of whether the balance between flexibility and accountability is optimal under the new arrangements will likely demand more attention in the coming years as the extent of progress towards outcomes becomes clearer.

Third, there are factors beyond the institutional features and processes of the Intergovernmental Agreement that are critical to the successful implementation of the benchmarking arrangements. Like all major public policy reform, the Intergovernmental Agreement challenges conventional practices. Many of the key features of the new framework require cultural change in the way all governments approach intergovernmental relations, policy development and service delivery — both across and within governments. The Council has called for greater cooperation and collaboration, trust and political leadership to support the reform agenda (COAG Reform Council 2010c, pp. 14–15).

References

- COAG (Council of Australian Governments) 2008a, *Communiqué*, 26 March http://www.coag.gov.au/coag_meeting_outcomes/2008-03-26/docs/communique20080326.pdf (accessed 26 October 2009).
- 2008b, *Intergovernmental Agreement on Federal Financial Relations*, http://www.coag.gov.au/intergov_agreements/federal_financial_relations/index.cfm (accessed 26 October 2009).
- 2008c, *National Education Agreement*, http://www.coag.gov.au/intergov_agreements/federal_financial_relations/docs/IGA_ScheduleF_national_education_agreement.pdf (accessed 26 October 2009).
- 2008d, *National Partnership Agreement on Literacy and Numeracy*, http://www.coag.gov.au/intergov_agreements/federal_financial_relations/docs/national_partnership/national_partnership_on_literacy_and_numeracy.pdf (accessed 26 October 2009).
- 2008e, *National Partnership Agreement to Deliver a Seamless National Economy*, http://www.federalfinancialrelations.gov.au/content/national_partnership_agreements/OT001/Seamless_National_Economy.pdf (accessed 3 November 2010).
- COAG Reform Council 2009a, *National Education Agreement: Baseline Performance Report for 2008*, COAG Reform Council, Sydney.
- 2009b, *National Partnership Agreement to Deliver a Seamless National Economy: Report on Performance 2008-09*, COAG Reform Council, Sydney.
- 2010a, *National Education Agreement: Performance Report for 2009*, COAG Reform Council, Sydney.
- 2010b, *National Partnership Agreement to Deliver a Seamless National Economy: Report on Performance 2009-10*, COAG Reform Council, Sydney.
- 2010c, *COAG Reform Agenda: Report on Progress 2010*, COAG Reform Council, Sydney.
- 2011a, *Seamless National Economy: Report on Performance*, COAG Reform Council, Sydney.
- 2011b, *National Partnership Agreement on Literacy and Numeracy: Performance report for 2010*, COAG Reform Council, Sydney.
- 2012a, *National Partnership Agreement on Literacy and Numeracy: Performance report for 2011*, COAG Reform Council, Sydney.

— 2012b, *Indigenous reform 2010–11: Comparing performance across Australia*, COAG Reform Council, Sydney.

Commonwealth of Australia 2009, *Australia's Federal Relations, Budget Paper No. 3 2009 10*, Canberra.

Commonwealth of Australia, Department of Finance and Deregulation 2012, *Is Less more? Towards better Commonwealth performance*, Discussion Paper, Commonwealth Financial Accountability Review.

Ekholm, M. 2004, 'Evidence-based policy research — some Swedish lessons', paper presented at the First OECD Conference on Evidence Based Policy Research, 19–20 April, Washington, <http://prod.ceg.rd.net/usermedia/images/uploads/PDFs/OECD-Ekholm.pdf> (accessed 2 June 2009).

MCEETYA (Ministerial Council for Education, Employment, Training and Youth Affairs) 2008, *2008 National Assessment Program — Literacy and Numeracy: Achievement in Reading, Writing, Language Conventions and Numeracy*, http://www.naplan.edu.au/verve/_resources/2ndStageNationalReport_18Dec_v2.pdf (accessed 27 October 2009).

12 Australian perspectives on benchmarking: the case of school education*

Peter Dawkins¹
Victoria University

12.1 Introduction

Improving the educational outcomes for children and young people is central to the nation's social and economic prosperity. In Australia, there are a complex set of Commonwealth–State relationships and division of powers in relation to school education (see Banks, Fenna and McDonald, this volume). This chapter focuses on school education in the Australian federal context and the major benchmarking developments that have occurred in this area over the last twenty years.

The Australian States have constitutional responsibility for school education, including the administration of government schools; development and delivery of curricula; and the regulatory conditions to ensure quality standards across all schools (including non-government schools).

Although State governments have primary responsibility for education, the Commonwealth has assumed an increasingly important role, driven to a significant extent by vertical fiscal imbalance. While States have the major service delivery responsibilities including in school education, they rely on substantial transfers from the Commonwealth, which raises the majority of the tax revenue. This resulted in a large number of specific purpose payments often with 'input controls'. One such example was the requirement for schools to have a flagpole carrying the Australia

* The author would like to acknowledge the support of Dr Sara Glover (Executive Director, Research and Analysis Division, Victorian Department of Education and Early Childhood Development), in the preparation of this chapter.

¹ Peter Dawkins was the former Secretary, Department of Education and Early Childhood Development, Victoria. He is currently Vice-Chancellor and President of Victoria University.

flag (Australian Government Programs for Schools, Quadrennial Administrative Guidelines 2005–2008).

Following the election of the Rudd Labor government in 2007, a process of reform was entered into, known as the National Productivity Agenda, following many of the ideas that had been proposed by Victorian Labor government in their proposed National Reform Agenda (DPC and DTF 2005). This involved the development of a national Education Agreement, involving payments from Commonwealth to State governments, linked with an outcomes framework; progress measures and targets; and an accountability and review framework, rather than detailed input controls. The National Education Agreement was supplemented with the National Partnership Agreements (on which, see Banks, Fenna and McDonald, this volume; O’Loughlin, this volume) to fund specific reforms and facilitate and/or reward States and Territories that deliver on these nationally significant reforms.

Thus benchmarking educational outcomes jurisdiction by jurisdiction became a key feature of the National Productivity Agenda, with a view to identifying those jurisdictions that implemented successful reforms that improved outcomes. This idea was along the lines of the former National Competition Policy. There is very little doubt that this agenda has motivated significant efforts in education systems to improve educational outcomes, and, in the authors’ view, has conceptual underpinnings as way of promoting educational progress in State systems in a world of vertical fiscal imbalance. However, it does bring with it a number of challenges that the Commonwealth and State governments have needed to confront in seeking to implement the policy successfully. This chapter discusses some of those challenges.

12.2 Background

There has been a strong history of benchmarking practices in school education across Australia. In 1993, the Heads of Government — now the Council of Australian Governments (COAG) — commissioned the *Report on Government Services* to help improve the effectiveness and efficiency of government services (see Banks and McDonald, this volume). COAG confirmed in late 2009 that the *Report on Government Services* should continue to be the key tool to measure and report on the productive efficiency and cost effectiveness of government services.

This framework of performance indicators aims to provide comparative information on the equity, efficiency and effectiveness of Commonwealth and State and Territory government services. The performance information promotes transparency and accountability; identifies areas of strong or poor performance; promotes

learning across governments; and creates an incentive to improve the performance of government services.

More recently, in 2008 the COAG Reform Council (CRC) was established to assist COAG drive its reform agenda by strengthening public accountability of the performance of governments through independent and evidence-based monitoring, assessment and reporting (see O’Loughlin, this volume).

The CRC reports on:

- the performance of the Commonwealth, States and Territories in achieving the outcomes and performance benchmarks specified in the National Agreements
- whether predetermined performance benchmarks have been achieved under National Partnerships.

In the case of National Partnerships, the CRC is the independent assessor of whether predetermined milestones and targets have been achieved. The assessment of each jurisdiction’s performance is reported publicly and a decision to make reward payments is based on this independent CRC assessment.

The latest development in benchmarking in School Education in Australia saw all Education Ministers agree to the publication of school information on the *My School* website. *My School* was launched in January 2010 by the Australian Curriculum and Assessment Authority (ACARA). ACARA was established by Commonwealth legislation in 2009, and is an independent authority responsible for the development of a national curriculum; a national assessment program; and a national data collection and reporting program that supports 21st century learning for all Australian students. ACARA has responsibility for publishing nationally comparable data on all — almost 10 000 — Australian schools on its *My School* website. Data on each school’s performance and factors relating to performance are provided. These include national testing in literacy and numeracy results; school attainment rates; student background characteristics; and information about each school’s teaching staff and income. Schools can compare their results in national literacy and numeracy tests with the results of other schools that serve similar students. They can also compare their progress against that of others schools that had the same starting point in 2008.

All results can be compared with results in statistically similar schools (that is, schools with similar student populations) across the nation. The *My School* website has been developed so that there is greater transparency and accountability about the performance of schools, and parents and the community have access to this information about their child’s school. The greatest potential value of this form of benchmarking lies in the support it provides for productive discussions between

schools that are doing better in similar circumstances to help them review and improve their own practices.

12.3 National Productivity Agenda and benchmarking school education

COAG agreements in late 2008 — the *Intergovernmental Agreement on Federal Financial Relations* and the *National Education Agreement* — resulted in all governments agreeing to a common framework for reform in education. An outcomes framework was developed and agreed to, establishing a set of aspirations, outcomes, progress measures and future policy directions to guide education reform in Australia — including a focus on improving outcomes for indigenous children and young people as well as students from low socio-economic backgrounds.

The underpinning idea of the National Productivity Agenda is that investment in human capital in the form of education and training raises work force participation and productivity, and therefore economic growth, and in turn government tax revenue (PC 2006; Dawkins 2010). This creates a virtuous circle.

In relation to school education, national and international evidence reviewed by the Productivity Agenda Working Group pointed to literacy and numeracy as a key determinant of school retention and subsequent workforce success, school completion as a major determinant of labour force participation, and teacher quality as the main in-school determinant of student success.

Using international benchmarking (especially using PISA data) it was also found that while the average performance of Australian students (for example, in literacy and numeracy at aged 15) is quite high by international standards, the performance of students from low socio-economic backgrounds was mediocre by international standards. Thus, key policy thrusts in the National Partnership Agreements included a focus on raising teaching quality, on raising literacy and numeracy and on improving outcome in schools with disproportionate numbers of students from low socio-economic backgrounds.

The very large gap in outcomes between indigenous and non-indigenous students was also a major focus. Consequently, the benchmarking that was agreed upon, was not only for average outcomes especially in literacy and numeracy and in school (year 12) completions, but also closing the gap between indigenous students and all students and between low-socio-economic background students.

12.4 What to benchmark

Determining *what* to benchmark is largely determined by the *purposes* of benchmarking. Benchmarking can be about promoting accountability and being transparent to the public about results. It can also be a practice that systematically evaluates relative strengths and weaknesses, searches for the evidence of improvements elsewhere, and more importantly how this was achieved (see Fenna, this volume). Benchmarking in the schools system aims to be both an assessment device and a learning tool.

Benchmarking in school education provides a mechanism of comparing the performance of the States and Territories. Currently these comparisons take different forms. These include the comparison of educational outcomes for children and young people; the comparison of inputs (for example, expenditure and teacher-student ratios); and the comparison of reforms and initiatives that have yielded improved results. Each of these different forms of benchmarking has their merits and value.

Benchmarking outcomes

The primary focus of the National Productivity Agenda is to benchmark outcomes because of the evidence that improving educational outcomes improves participation and productivity. Benchmarking outcomes compares performance between jurisdictions or cohorts on the actual results achieved by children and young people in each jurisdiction.

An example of this is benchmarking student reading results on the National Assessment Program — Literacy and Numeracy (NAPLAN) (figure 12.1).

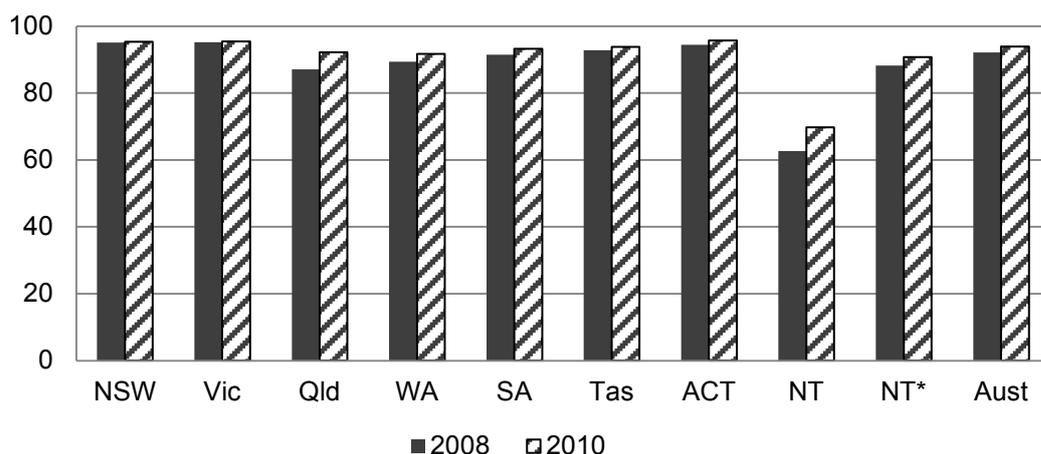
At face value, there is little difference between jurisdictions in these Year 3 reading results apart from the results for Indigenous students in the Northern Territory. Furthermore, with less than 10 per cent of children in Year 3 below the National Minimum Standard in reading one might question the value of this level of comparison. Other comparisons can also be made such as the proportion of students in the top or lower levels of performance and improvements of these over time.

Similarly comparison of the performance of different cohorts of students can be made. These include indigenous students; male and female students; students living in remote, very remote and metropolitan regions; students with language backgrounds other than English; and socio-economic status.

Most importantly we must be clear about the policy objectives and the outcomes we are trying to achieve and benchmark accordingly. Benchmarking whole year levels will mask differences and gaps in performance between cohorts of students and the

distribution of outcomes at the student and school level. It is possible, for instance, that improvements in student outcomes may occur at a jurisdictional level, yet not be apparent among low SES students.

Figure 12.1 **Year 3 reading: proportion of students at or above the national minimum standard**



* 'NT' includes all Australians resident in the NT. 'NT*' includes only non-Indigenous Australians resident in the NT.

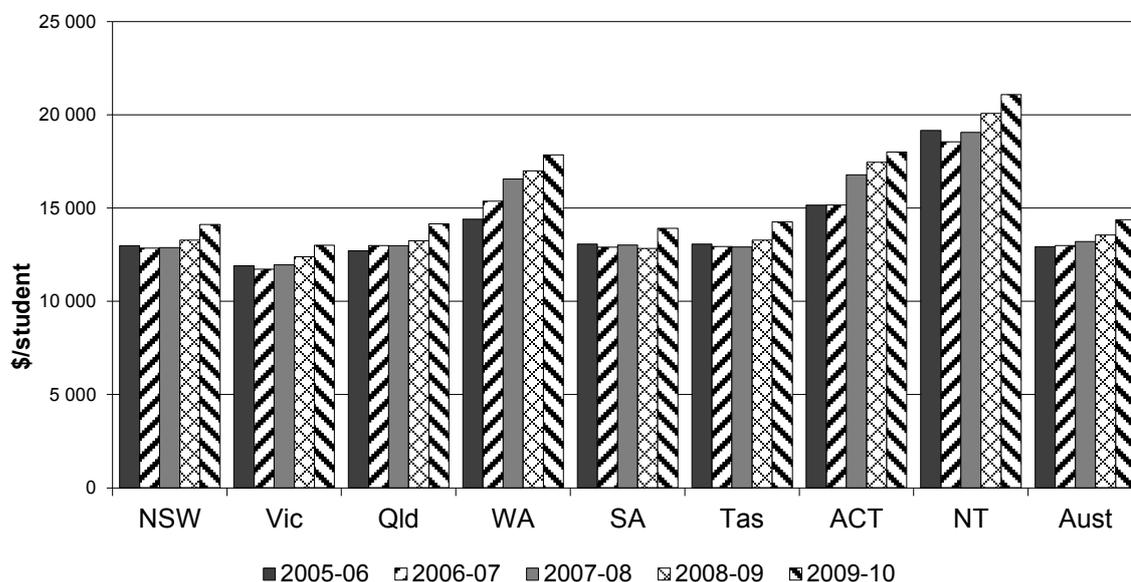
Source: COAG Reform Council (2011).

However, once we start benchmarking for sub-populations of students, issues of measurement error emerge. If this type of outcomes-benchmarking is used for evidence-informed decision-making then it is more likely that margins of error can be tolerated. Once external publications, performance judgments, and rewards or sanctions are applied to these types of benchmarking activities, there is potential to undermine or limit the ambition of the reform and/or the target group to be measured (Fenna, this volume). Attention and debate turn to issues of data and measurement, rather than policy and strategy. We return to this important issue in benchmarking later in this chapter.

Benchmarking inputs–outputs

Perhaps the most longstanding benchmarking practice in school education has been on benchmarking inputs and outputs. Examples include the recurrent expenditure per full-time equivalent student and student–teacher ratios.

Figure 12.2 **Government real recurrent expenditure per full time equivalent student, government schools (2009-10 dollars)**



Source: Steering Committee for the Review of Government Service Provision (2012).

The variation on government expenditure between jurisdictions is evident. What conclusions can be drawn from these benchmarked data? Does higher expenditure mean a less efficient system, or might it mean simply that the cost of educating students in more remote areas of Australia is higher and vice-versa does lower expenditure mean a more efficient system and greater economies of scale can be achieved in more densely populated areas? There is the potential for quite misleading conclusions to be drawn.

Similar issues arise with Student–Teacher ratios and average class size data. There is evidence that government investment in reducing class size is less effective than investments in improving teacher quality (Jensen 2010).

Such benchmarking creates a public pressure to maintain low class sizes without necessarily improving the quality of teaching and the outcomes for students. Furthermore, in jurisdictions with devolved decision-making regarding hiring of staff, schools may decide to employ a broader range of non-teaching staff to support individual students and families and provide necessary support for teachers. Benchmarked data on teacher numbers alone will not take these variations into account.

Benchmarking reforms

A question for policy makers is to understand how school systems improve performance. This type of benchmarking analyses the reform elements or sets of interventions of different systems that have led to significant gains in student outcomes as measured by national and international assessments. McKinsey & Co. (2007), for example, have undertaken international comparisons of school systems and subsequently Barber (2009) developed an empirically based framework for assessing the progress of different systems on a number of key dimensions.

Table 12.1 **Benchmarking system reform: nine characteristics**

<i>Standards and Accountability</i>	<i>Human Capital</i>	<i>Structure and Organisation</i>
<ul style="list-style-type: none"> • Globally-benchmarked standards • Good, transparent data • Every child on the agenda always in order to challenge inequality 	<ul style="list-style-type: none"> • Recruit great people and train them well • Continuous improvement of pedagogical skills and knowledge • Great leadership at school level 	<ul style="list-style-type: none"> • Effective, enabling central department and agencies • Capacity to manage change and engage communities at every level • Operational responsibility and budgets significantly devolved to school level

Source: Barber 2009.

Barber's framework incorporates three key themes which in his analysis have been the key to successful school systems. First there need to be rigorous performance standards against which schools and their students are to be assessed, and there needs to be appropriate levels of accountability for that performance. However, this cannot be successful without the second theme, which Barber calls the human capital in the system. That is, there needs to be a capacity building agenda to build an effective school workforce to be able to achieve the performance standards. Third, school systems need to be structured on a way that takes advantage of being a system, whole devolving appropriate responsibility to the school level. Under each of the three themes Barber identifies three characteristics that need to be present. This provides a basis for a school system to assess itself against what Barber concludes from his analysis is best practice for a school system.

Furthermore it is possible to benchmark the delivery of reforms or *how* a system implements the interventions. By doing this, systems can evaluate the relative strengths and weaknesses of the policy-implementation or delivery chain.

While these different benchmarking practices in school education in Australia offer important information for the public and for policy makers, the interplay between transparency, accountability and improvement in different contexts create a number of major challenges that may have both intended and unintended consequences.

12.5 Issues in benchmarking

The foregoing analysis has identified a very important role for benchmarking in seeking to improve educational outcomes for school students. However, there are a wide range of challenges in undertaking successful benchmarking. Furthermore, when the stakes are raised and when benchmarking itself becomes an integral part of the incentive and reward system — as under the National Partnerships — these challenges can make it difficult for the policy to be implemented successfully.

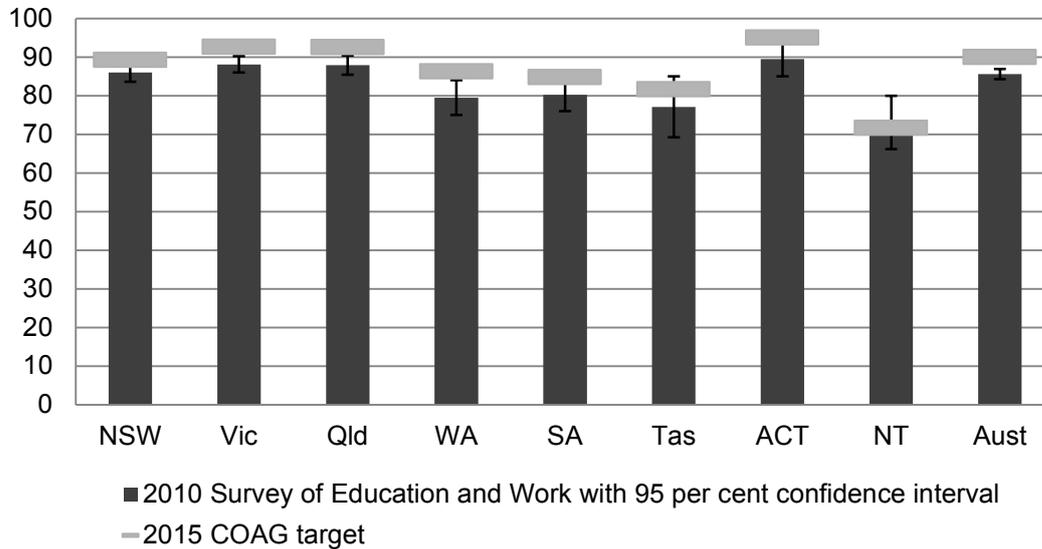
In this section we identify some of these challenges especially in relation the problems that have arisen under the National Education Agreement.

Measurement

Benchmarking for National Partnerships and My School requires nationally consistent and comparable data. The National Assessment Program — Literacy and Numeracy (NAPLAN), an annual national test for students in Years 3, 5, 7, and 9, is used for benchmarking literacy and numeracy. There are limitations of measurement in this assessment of students at the very top and bottom ends of performance. There is also measurement error. When it comes to ‘judging’ improved performance for sub-populations of students as well as comparing the performance of schools, the problem of measurement error cannot be discounted.

Similarly the only national comparable data for school attainment (year 12 or equivalent certification) is a national survey of education and work conducted by the Australian Bureau of Statistics. This is an annual sample survey and useful at a national level. However, when results are disaggregated by jurisdiction, large confidence intervals emerge. These make it difficult to compare jurisdictions; impossible to assess whether results in jurisdictions are improving over time; and relatively meaningless in trying to understand what is happening to Indigenous students, students from low socio-economic backgrounds and students in regional and rural locations in different States and Territories.

Figure 12.3 20–24 year olds with year 12 or equivalent or certificate II, per cent



Source: ABS (2011), Survey of Education and Work 2010.

Context

When benchmarking educational outcomes, it is important to take into account the different contexts. Thus, students from lower socio-economic backgrounds, for example, tend to underperform relative to students from higher socio-economic backgrounds. On the one hand, these effects need to be taken into account when benchmarking, and rewarding schools and school systems. On the other hand, one of the aims of school improvement is to reduce those effects, so benchmarking should not assume that these affects are set in stone.

Levels versus changes

When benchmarking measures of, for example, literacy and numeracy, it is important — partly because of the different contexts mentioned above — to look both at levels of student performance and changes in them. Thus it would have been unreasonable under the NAPLAN testing regime to expect that students in the Northern Territory will perform as well as the students from the ACT. Thus, in providing incentives and rewards the focus has been on making improvements in the levels. However, this raises further questions. For example, is it reasonable to expect students from the Northern Territory to improve faster than students from the ACT or vice versa?

When examining the trajectory of schools on the *MySchool Website*, if a school is already a high achiever, especially taking into account their socio-economic mix, how serious is it if in a particular year they appear to deteriorate? Should schools with very poor performance, or systems containing those schools, that improve marginally, be rewarded for so doing?

In putting forward targets for rewards, if a system is showing declining literacy or numeracy, should they be rewarded simply for slowing the decline, or should they actually achieve a real improvement?

These are very difficult questions to which there is no simple answer.

What outcomes to focus on

The main foci of the National Education Agreement have been literacy, numeracy and year 12 attainment. All these outcomes can be strongly defended as important and it is very good when outcomes can be improved in these areas.

Having said that, there is a very legitimate argument that there are other objectives of school education. Indeed the *Melbourne Declaration on Educational Goals for Young Australians* identified an array of outcomes that we should seek to focus on under the general headings of ‘Successful Learners’; ‘Confident and Creative Individuals’; and ‘Active and Informed Citizens’. Indeed there is a growing literature on 21st century skills that will be required by students graduating from our education systems (see for example Wagner). In general, improved literacy and numeracy are likely to be a significant input into many of these skills. On the other hand, if schools focus overwhelmingly on improving measurable outcomes in literacy and numeracy, and pay scant attention to other areas such as teamwork, problem solving, cross-cultural and communication skills, this may not be to the benefit of the students. As Fenna (this volume) notes, ‘teaching to the test’ (and ‘neglecting the broader suite of often less tangible or immediate desiderata’) is a perennial risk in performance monitoring regimes.

It is important to develop useful measures of these other attributes and use them in improving education and benchmarking schools and systems, but similar issues will arise about what weights to apply to the different measures.

Causality

When seeking to improve educational outcomes with the assistance of benchmarking, it is important to develop an understanding of what causes

improvements in outcomes. This requires sophisticated and in-depth analysis and evaluation. This can include multivariate statistical analysis, but also appropriate forms to share information about ideas and best practice.

When using targets to reward performance, the reward will flow whether the cause is anything to do with change in policy or practice. This may not be a problem provided it does create the right incentives and policy and practice does improve one way or another. However, it is important to undertake evaluation of the targets and their effects and refine and improve them over time.

Timeframes and orders of magnitude for expected improvement

To date, it is unclear what a reasonable timeframe for system improvement is. There has been little benchmarking on actual improvement taking into account different starting points. For a system to improve a mean score or lift a proportion of students above a particular standard takes time, and the evidence linking interventions to improvement over time remains limited.

Furthermore, improvements over time can be misleading. A feature of NAPLAN results is that improvements of systems, schools and students with lower starting points are likely to be greater than improvements observed from higher starting points. Unless some form of relative improvement is taken into account, absolute improvement scores can be quite misleading.

Hattie suggests that benchmarking in schooling should focus attention on the growth in student learning. For this to occur, education systems need to provide the tools and incentives to monitor individual student's progress, rather than relying on standardised scores and minimum standards.

Commonwealth–State relations and the use of benchmarking

The use of benchmarking to provide incentives and rewards to State and Territory jurisdictions for improved educational outcomes was a feature of the National Reform Agenda proposed by Victoria, which was ultimately adopted by COAG on the National Productivity Agenda. In Victoria's proposal the outcomes and measures would be developed and implemented in a collaborative federalist model, and administered by a federal entity reporting to COAG. In practice the Commonwealth Government who provided the funds for the incentives and rewards, itself runs the incentive and reward mechanisms albeit with some consultation and support from the COAG Reform Council, which provides periodic reports on the progress of systems in improving educational outcomes. This

approach brings with it the risk that it can become a top-down coercive performance monitoring model, rather than a collegial model which is more genuinely cooperative in nature with mutual goal development and standard setting, which would perhaps be more focussed on learning and improvement.

12.6 Conclusion

This chapter has canvassed a wide range of issues relating to benchmarking in school education. In a world of increasing transparency and accountability, good measurement and successful benchmarking, is a critical aspect of good educational policy. In Australia, through the National Education Agreement, and related developments, such as national testing of literacy and numeracy, benchmarking has become firmly embedded in national policy through the setting of incentives and rewards for state and territory jurisdictions.

This represents a very significant development in evidence based policy in Australia, and in the authors' view, continues to have great potential as a way of managing effective Commonwealth–State relations in a world of shared responsibility for education and vertical fiscal imbalance.

Nonetheless it brings with it significant challenges with which Australian policy makers are grappling. The problems are unlikely ever to be 'solved' in a permanent way — like most policy problems. But over time it is to be expected that we will learn from experience and make significant progress in the bid to promote educational outcomes in Australia. It is very important that the use of data and benchmarking to create incentives and rewards, is only one part of a broader approach to school improvement, that keeps schools, school systems and governments, focused on the essence of improvement, rather than just on measurement issues alone.

References

- Barber, M. 2009, *School Leadership and Whole System Reform*, presentation to EU Education Ministers Event, Sweden, 24 September 2009.
- COAG Reform Council 2011, *Education 2010: Comparing performance across Australia*, Statistical supplement, Report to the Council of Australian Governments, Sydney, September.
- Dawkins, P. 2010, 'Institutionalising an evidenced-based approach to policy making: the case of the human capital reform agenda', Chapter 11 in *Strengthening Evidence-Based Policy in the Federation*, Roundtable

Proceedings, Productivity Commission, Canberra, 17-18 August 2009, Volume 1.

DPC and DTF (Department of Premier and Cabinet and Department of Treasury and Finance) 2005, *Governments Working Together: A Third Wave of National Reform – A New National Reform Initiative for COAG: The Proposals of the Victorian Premier*, DPC, Melbourne, Victoria.

Hattie, J. 2009, *Visible Learning: a synthesis of over 800 meta-analyses relating to achievement*, Routledge, London.

Jensen, B 2010, *Investing in Our Teachers, Investing in Our Economy*, Grattan Institute, Melbourne.

McKinsey & Co. 2007, *How the World's Best Performing School Systems Come out on Top*, http://www.mckinsey.com/clientservice/socialsector/resurces/pdf/Worlds_School_Systems_Final.pdf.

PC (Productivity Commission) 2006, *Potential Benefits of the National Reform Agenda*, Report to the Council of Australian Governments, Research Paper, Productivity Commission, Canberra.

Steering Committee for the Review of Government Service Provision 2012, *Report on Government Services 2012*, Productivity Commission, Canberra.

Wagner, T., 2008 *The Global Achievement Gap: Why Even Our Best Schools Don't Teach the New Survival Skills Our Children Need – And What We Can Do About It*, Basic Books, New York.

13 Benchmarking in federal systems: the Queensland experience

Sharon Bailey¹

Queensland Department of the Premier and Cabinet

Ken Smith²

Queensland Department of the Premier and Cabinet

During the federation debates in the 1890s, Queensland's contribution to the constitution, through Sir Samuel Griffith, was primarily to ensure the allocation of residual powers to the States and the concept of equal State representation in the Senate. This focus on the rights of the States has been maintained since, with Queensland sometimes forming alliances with other States and Territories, across party political lines, to strengthen its bargaining power and maintain the federal balance.

As chapter 8 details, today's federation is quite different from where Australia started over a century ago (Banks, Fenna and McDonald, this volume). The progressive erosion of the State revenue base, the expansion of the Commonwealth's power, and a lack of extra-constitutional mechanisms to allow for formal collaboration and joint decision-making, has meant that we have transitioned from a coordinate federation to one of policy interdependence and overlap (Fenna 2007). Within this context, benchmarking has a special place. It has become an important part of the contractual process by which funding arrangements are managed and comparisons between and within the other States and Territories made. It has become shorthand for the very complex area of performance assessment.

This chapter details four benchmarking exercises that Queensland is involved in with the Commonwealth. These case studies are as follows:

- Elective Surgery

¹ Sharon Bailey is the Executive Director, Office of the Director-General, Queensland Department of the Premier and Cabinet.

² Ken Smith is the former Director General of the Queensland Department of the Premier and Cabinet.

-
- National Assessment Program — Literacy and Numeracy (NAPLAN)
 - The Remote Indigenous Housing Agreement
 - Cape York Welfare Reform.

These encompass a variety of arrangements, each with their own unique histories and sets of relationships and provide a sample of the sundry ways that benchmarking can be used and the diversity of its impacts, on both services, and the relationships between jurisdictions within a federal system.

13.1 Benchmarking

As Fenna (this volume) notes, the term ‘benchmarking’ is used loosely to cover measurement regimes that have a comparator or target and provide:

- accountability for taxpayer dollars through greater transparency
- improvements to services and ultimately the quality of life enjoyed by individuals, families, targeted groups and, hopefully, the whole community.

Given the effort, time and cost involved in setting up data collection and benchmarking systems, it is important that we do not become distracted from these two fundamental goals.

It is a truism to say ‘what gets measured gets done’, but equally, ‘what gets measured *regularly*, gets done *habitually*’. Hence, it is extremely important to select the areas of measurement carefully, so that energy is focused on the priority areas — not just the things that are easy to measure, which may well not be as important, and may take energy away from those things that are.

The decisions we make about what to measure and report on, have a direct and significant impact on the behaviour of front-line staff, even before the reports come in. To ensure that the impact works in the interests of the broader community, we have to think carefully about how the design, analysis, distribution and use of that information will help people do their job better.

Additionally, it is important to acknowledge the limits of benchmarking. There is a persistent myth that somehow science and evidence can simplify our decisions and solve our problems. And this myth persists, despite our experience. To quote Donald Schön:

There is a high, hard ground where practitioners can make effective use of research based theory and technique and then there is a swampy lowland where situations are confusing messes incapable of technical solution. The difficulty is that the problems of

the high ground ... are often relatively unimportant to clients or to the large society, while in the swamp are the problems of greatest human concern. (Schön 1983, pp. 42–3)

Benchmarking and performance measures are extremely useful additions to the repertoire of policy tools, but they are the beginning of a conversation, not the final word. They raise questions that need to be investigated in pursuit of improving what we offer to the community; they rarely, in and of themselves, provide the answers. It's that next step that often appears to be missing.

13.2 Elective surgery

Within the Australian federation, health services are delivered by a variety of government and non-government providers. There is a significant overlap between the Commonwealth and the States, which has been the subject of the recent Health Reform process.

Public hospitals are funded by both levels of government. The Commonwealth currently funds approximately 35 per cent of Queensland's public hospital services, with the bulk of the remainder being made up by the State and a small percentage provided by other sources such as health insurance funds and workers compensation. The administration and delivery of public hospital services, however, is a State responsibility.

When people think about measures for health and hospitals, elective surgery waiting lists are often top of mind. While these may only be a second tier indicator, elective surgery waiting lists are critical to the public perception of the overall effectiveness and efficiency of the health system. Indeed, the intensity of public feeling has led to debates in the media about the potential for the Commonwealth to take over the full management of public hospitals. Hence, there is significant focus from both levels of government on ensuring that waiting lists are kept down.

Measurements in regard to elective surgery — ostensibly the time from when patients are added to a waiting list to the date on which they are admitted, classified into clinical urgency categories — have been reported on for many years. Currently they are the subject of the *National Partnership Agreement on the Elective Surgery Waiting List Reduction Plan* between the Commonwealth and States and Territories to be reported to and on by the COAG Reform Council (on which, see O'Loughlin, this volume). But prior to the Agreement, Queensland and other jurisdictions have been contributing data on elective surgery voluntarily to the *Report on Government Services* (RoGS) for over a decade (on which, see Banks and McDonald, this volume). Additionally, Queensland Health publishes quarterly hospital performance

reports that include elective surgery wait times on its website, along with a specific report that focuses on the quarterly performance of Queensland against the Partnership Agreement.

Elective Surgery wait times were a particular feature of the major *Queensland Health Systems Review* in 2005, sparked by the *Patel Inquiry* — an Inquiry into a health practitioner’s clinical outcomes at the regional Bundaberg Hospital. Queensland was found not to be meeting the established benchmarks — not only in comparison to other States, but more importantly, in relation to the clinically recommended wait periods for particular surgical categories.

Public Hospital/Health System crises have the capacity to galvanise political will. In response, significant resources were redirected to deal with this issue. In addition to *Surgery Connect*, there was also significant business process reengineering and as a result, Queensland now has a much more streamlined process for patients and a much more effective and efficient use of surgery theatres across the State.

Queensland is now performing well against the key indicators. the State government has allocated significant financial and human resources to reducing wait times including initiating the *Surgery Connect* program whereby Queensland Health has paid for public patients to have their operation in the private system, as a means of clearing some of the backlog, and increasing system capacity/throughput. In itself, *Surgery Connect* provides a basis for benchmarking the costs and effectiveness of public compared to private provision.

This is a very positive benchmarking story. Benchmarking helped highlight a system deficiency that was affecting quality of life; improvements were made; performance against the benchmark improved; people are now getting their surgery within clinically recommended times; and Queensland, along with most other States, has received a reward payment under the Partnership Agreement. All in all, it has been a win–win situation. That said, it is important that other factors are considered.

The RoGS report provides comparative data across States and across time-series on elective surgery waiting times for clinical urgency categories 1, 2 & 3. However, different States include different things in their categories, and, of course, it is in the interest of the State to include as little as possible in Category 1, as that has the shortest time frame. Hence, it could be argued that the very act of reporting begins to influence behaviour. Consequently, comparisons across jurisdictions are often not valid. The Productivity Commission is extremely clear about this in its report, but once something is in a table the subtleties are often lost, and a judgement is made regardless.

Secondly, one needs to understand the context and history of waiting lists. Being added to a public waiting list for elective surgery is a process in and of itself. During the *Queensland Health Systems Review*, it was discovered that there were waiting lists behind the waiting lists — that is, to get onto an elective surgery waiting list it was necessary to see a surgical specialist, but there were long waiting lists to see the surgical specialists. The people on the surgical specialist waiting lists were technically waiting for elective surgery, but they were not showing up on the official elective surgery lists, as they did not meet the technical precondition for that list.

Additionally, at that time Queensland Health regions offered financial incentives to hospitals on the basis of a 5 per cent long wait performance benchmark, that led to some gaming of the system. Those loopholes were closed, but any system can be gamed. The people gaming the system in this case were doing so to try and maximise the operation of their hospital in an environment of resource constraint. The purpose of the gaming was to procure necessary resources for the whole of the patient population, but the end result was that the publicly available reporting was not accurate.

Finally, elective surgery waiting lists are a second tier indicator. To quote Peter Forster, who undertook the Health Review in 2005:

The current community and media focus on elective surgery waiting lists whilst understandable at one level, is not the best overall indicator of health service performance nor is it necessarily in the best interests of all patients. Waiting lists are an imprecise indicator of the level of access to public hospital services and place undue focus on certain kinds of surgical activity sometimes to the detriment of medical services. Due to budget and workforce constraints the community's need is not being met which is resulting in less than optimal patient outcomes. Surgical waiting lists reflect Queensland Health's attempts to manage finite resources where demand for services exceed supply. (Forster 2005, p. 122)

This raises the question of the extent of the opportunity costs associated with such a focus on elective surgery. Does such an intense focus come at the detriment of other more important facets of the health system? It is important that these questions remain at the forefront of our efforts, so that we drive whole-of-system improvement.

13.3 National Assessment Program — Literacy and Numeracy (NAPLAN)

Australian State and Territory governments have responsibility to ensure the delivery of schooling to all school-age children and provide the bulk of the funding

for that provision. The Commonwealth provides supplementary funding for government schools through the National Education Agreement (NEA), and for non-government schools through the *Schools Assistance Act 2008*. Additionally, the Commonwealth Government has been working with States and Territories to implement a National Curriculum. Like health, education not only represents a large part of government spending, it is an area of intense public focus.

The NAPLAN test is an annual, census-style test that was first administered across Australia in 2008. Results are reported for each of the domains of Reading, Writing, Spelling, Grammar and Punctuation, and Numeracy, with six bands of achievement being used for reporting student performance in each year level (Years 3, 5, 7 and 9).

There are three performance measures used to describe NAPLAN results:

- National Minimum Standard (NMS) — which represents the attainment of only the basic elements of literacy and numeracy for the year level
- Mean (Average) Scale Score (MSS)
- Upper Two Bands (U2B) — which shows the proportion of students achieving in the upper two bands for each year level.

NAPLAN occurs in the context of a National Partnership Agreement on Literacy and Numeracy, which has a budget of \$540 million with an additional \$30 million allocated to fund Literacy and Numeracy pilots in low SES communities. This Agreement operates for four years from 2009 and contains both facilitation and reward payments. Reward payments are dependent on evidence of literacy and numeracy progress and achievement, monitored through:

- NAPLAN results — Years 3, 5 and 7
- progress on P-9 Literacy and Numeracy indicators
- validated teacher judgements through formal assessments (Assessment Bank) and annotated samples of student work
- progress on ESL Bandscale for students from non-English speaking background.

Queensland's results in the initial NAPLAN testing were disappointing, and consequently the Premier commissioned an independent study into Primary Schooling, by Professor Geoff Masters of the Australian Council of Educational Research, with a view to: a) testing whether there really was a problem; and b) if there was, finding a way to address it.

The Masters' Review concluded that there was indeed a problem, although it cautioned against drawing inferences about the quality of education in Queensland

based solely on comparisons of Queensland mean achievement with that of other States and Territories in NAPLAN and other tests.

The review made five recommendations:

- tests for aspiring teachers to demonstrate threshold knowledge in teaching literacy, numeracy and science
- a new structure and program of advanced professional learning for primary teachers
- additional funding for specialist literacy, numeracy and science teachers in districts/schools where they are most needed
- standard science tests in Years 4, 6, 8 and 10
- an expert review of school leadership with a view to establishing a program of professional learning for primary school leaders to drive improved performance.

These recommendations were largely adopted in full, and again, this is a positive story. The testing uncovered a problem that Queensland had suspected, but NAPLAN gave it a form and gave the impetus to address it in a concerted way. The ensuing review was of a high quality and recommended five substantial, fundamental system improvements. In addition, further analysis of the NAPLAN results allowed the Department of Education and Training to drill down and uncover key problem areas. For example, in literacy, students in Years 3 and 5 were struggling with figurative thinking, use of pronouns and sentence structure, so particular remedial programs on those areas were able to be designed.

Queensland has made significant progress in implementing those system improvements, and students are benefiting as a result.

- Nine out of ten Queensland students are meeting the NMS for literacy and numeracy, with the strongest result in Year 3 Numeracy at 95.2 per cent of NMS, and the weakest result being Year 9 Writing at 84.7 per cent at NMS (noting the national average of 84.6 per cent).
- Queensland 2011 Year 3 students are the first full cohort to have passed through the Prep year, and have posted the State's strongest Year 3 results since NAPLAN testing began in 2008. Year 3 students have improved in all test strands for NMS, MSS and U2B as measured from 2008 to 2011 and from 2010 to 2011. This places the Queensland Year 3 students at fourth in the country for Reading and Grammar and Punctuation, and sixth for Spelling, Writing and Numeracy.
- Two cohorts have now sat NAPLAN twice, in 2009 and 2011. The Year 3, Year 5 and Year 7 students from 2009 were in Year 5, Year 7 and Year 9 in

2011. The gains made by Queensland students, from Year 3 – 5, Year 5 – 7 and Year 7 – 9, have exceeded the gains made by their counterparts in Australia overall in eleven of the twelve comparable test areas.

- Since 2010, Queensland has improved in nine of the sixteen comparable test strands for NMS; eight of sixteen strands for MSS; and eight of sixteen in U2B.
- Since 2008, Queensland has improved in fifteen of the sixteen strands for both NMS and MMS and thirteen strands for U2B; in most domains the difference between Queensland and the highest performing jurisdictions is only a few percentiles.

While Queensland's results have noticeably improved since 2008, the State's relative position has not altered to any great extent. It should be noted Queensland remains sixth across jurisdictions for average National Minimum Standard (NMS) scores and improved from seventh to sixth for Mean Scale Scores (MSS). One could say that this is due to the time lag between bedding down the improvements and improved performance, but equally for Queensland to improve dramatically in comparison to other States, requires either inertia or decline in the performance of the other States and Territories, which is perhaps not an appropriate ambition.

Additionally, the NAPLAN test was developed through negotiations between all the jurisdictions, in the absence of a National Curriculum. The implementation of a National Curriculum will play an important part in improving the consistency of inputs and hopefully the outcomes, across Australia.

There is also a question about the purpose of the testing regime itself: is this a test to assess the health of a jurisdiction's system or is it a diagnostic tool to assist students' capacity? If it's the former, random sampling would be a much more efficient methodology, but if it is the latter, we need to get much better at using it to understand how to improve teaching and learning results for individual students. Is it a test *of* learning or a test *for* learning? This has been discussed at length, but we are not clear. Benchmarks and measures cannot be all things to all people, but when there is a dearth of information, there is a tendency for them to be used in that way.

There is a vacuum in regard to information about the performance of Australian children's schooling and as a result, information like NAPLAN is seized upon. This can be seen in the overwhelmingly positive reaction of parents and the community to the *MySchool* website, which publishes NAPLAN and other data on all Australian schools.

Additionally, when there is an information vacuum, the little information that does exist, can be given disproportionate weight and influence.

The government has received numerous letters from parents reporting that their children have been asked to stay home on test day, and that NAPLAN results are being used by private schools to screen students applications. This is concerning and not what was intended by the people who designed the test.

This is not an argument against benchmarking; rather, it reminds us that data and reporting regimes can be misused, once they are up and running.

Finally, benchmarking is generally about supply, and yet particularly in the education area, one of the biggest indicators of success is demand. Unless you have demand, improvements to supply can be wasted. Benchmarking places the emphasis on the supply side, rather than looking at the preconditions for creating demand — a much more fundamental question.

13.4 Remote Indigenous Housing National Partnership Agreement

The *Remote Indigenous Housing National Partnership Agreement* (RIHNPA) was negotiated between the Commonwealth and State and Territory Governments to reduce severe overcrowding in remote Indigenous communities; increase supply of new houses and improving condition of existing houses; and ensure rental houses are well maintained and managed.

The RIHNPA provides Queensland with \$1.16 billion over 10 years (from 2008-09) to provide 1141 new dwellings and 1216 upgrades to existing social housing in remote Indigenous communities. Some of these areas are more than twice the distance between Brisbane and Melbourne away from the Capital, generally require four wheel drive vehicles, barges and aircraft to access, and can be cut off from surrounding communities for weeks or months during the wet season.

The Queensland government is providing \$32.4 million over five years to establish the Remote Indigenous Land and Infrastructure Program Office (the PO) which has responsibility for land and infrastructure planning issues across Queensland's remote Indigenous communities and the negotiation of and roll out of lease agreements. Queensland is also spending an additional \$67 million to address that backlog of infrastructure requirements in these communities.

Such a transition, to direct leasing by Government of communal lands for social housing purposes, is a sensitive and contentious issue for land-holders in remote communities.

The need to negotiate Indigenous Land Use Agreements (ILUA) where Native Title (NT) has not been extinguished, which is the case for most communities, has had a significant impact on the delivery of this Agreement, as has the Commonwealth's late withdrawal of a proposed Municipal Infrastructure National Partnership Agreement. But most importantly there was a late, complicating, element—the Commonwealth condition that a minimum 40-year lease was required before any new houses could be constructed. Queensland needed to secure these leases with individual Councils to protect capital investment, as there is no freehold land available in these areas (unlike in some other jurisdictions).

Negotiations with councils to obtain 40-year lease agreements took much longer than expected, with the first agreement to grant a lease being obtained in late February 2010. Negotiations were conducted in good faith with Councils and as a result, at the end of the 2009-10 financial year, seven of the fourteen eligible Aboriginal Shire Councils had signed Deeds of Agreement to Lease and Deeds of Agreement to Construct. (All have now agreed to leases for social housing.) This is a major achievement in normalising social housing arrangements.

The targets for the 2009-10 RIHNPA were 65 new construction/replacement houses and 150 upgrades. Queensland achieved 46 new construction/replacement houses and 152 upgrades. This was a significant achievement, given that construction could not commence until late February 2010.

Regardless, the Commonwealth advised Queensland in July 2010, that the State had been penalised 2.5 per cent or \$3.12 million, for not completing the 2009-10 targets on time. This amount was to be taken from Queensland's Employment Related Accommodation (hostel style or rental accommodation for people moving from remote Indigenous communities for work or training opportunities) funding for 2010-11.

Of course, Queensland would have preferred not to be penalised, and this was conveyed politely but firmly in writing, noting the impact of the late imposition of the 40-year lease condition. The response from the Commonwealth has indicated that Queensland was lucky not to have been punished more severely. Given that the delay was not due to recalcitrance or incompetence the benefit of the penalty can be called into question. It reinforced a long-held Queensland view that Canberra has no idea of the practicalities of delivering in remote communities.

Additionally, one could argue that the RIHNPA contained some incompatible targets; for example, local Indigenous employment targets and housing completion targets. Both are important targets in achieving the longer-term outcomes of the RIHNPA, but if you are seeking to complete housing targets within a tight time

frame, without a ready supply of local skilled labour, then there are likely to be delays.

In hindsight, Queensland should have renegotiated the targets for construction and upgrades following agreement with Councils, but renegotiating timeframes on agreements of this nature is politically unpalatable.

This case study brings the Commonwealth–State relationship into sharp relief, and raises the issue of sanctions. The three things that characterise a good contract are: information; certainty; and rewards/sanctions. Sanctions are critical, but have an impact on relationships and hence on performance going forward. In those cases where the sanction itself may have an adverse effect on future performance — that is, if financial resources are necessary to performing against the next stage of the contract — how does a financial penalty help that next stage of performance? State Budgets are large enough to make up the penalty; however, States have a whole range of other priorities, and the NPAs generally represented Commonwealth priorities. This is a problem that has dogged performance management systems since their inception and there are no easy solutions. However, it does highlight the question as to what motivates performance and what prevents non-performance, and the fact that sanctions are a strong, but potentially blunt, tool that requires supplementation.

13.5 Cape York Welfare Reform

The Cape York Welfare Reform trial is a very different exercise to the previous three examples. Benchmarking in this instance provides the data by which the Commonwealth and State can together evaluate a very new approach to welfare provision.

The Commonwealth and Queensland governments entered into a partnership with the Cape York Institute to deliver the Cape York Welfare Reform (CYWR) trial at the beginning of 2008. The trial will run for four years in four communities, affecting around 1800 people.

Its objectives are ambitious:

- restore positive social norms
- re-establish local Indigenous authority
- support community and individual engagement in the real economy
- give people choices around moving from social housing into home ownership.

It began in May 2007 with the release of the Cape York Institute for Policy and Leadership (the Institute), led by Mr Noel Pearson, *From Hand Out To Hand Up* report which proposed a ‘welfare reform trial’ in four communities — Hope Vale; Aurukun; Mossman Gorge; and Coen (the welfare reform communities). The Institute itself is a model of Commonwealth–State cooperation, half funded by the Commonwealth and half by Queensland.

Following this, the Federal Parliament amended social security legislation to enable the Commonwealth’s income management interventions in the Northern Territory; a national income management regime to apply in cases of child safety and school enrolment and attendance; and the proposed Cape York trial by anticipating the establishment of a ‘Queensland Commission’ to direct Centrelink to place a person under compulsory income management. It also provided exemption from the operation of anti-discrimination legislation and “special measure” status for this Commission and the Northern Territory intervention as these initiatives have an Indigenous focus.

The Family Responsibilities Commission (FRC) was then established under an Act of the Queensland Parliament, to directly link improved care of children to receipt of welfare and other government assistance payments. The FRC also connects families with support services to strengthen family roles. To do this, the FRC relies on notifications from Queensland government departments for breaches of State laws.

Bringing this into being has required an active partnership between the Commonwealth, the Queensland government, the Cape York Institute for Policy and Leadership (CYI) and the communities of Aurukun, Hope Vale, Coen and Mossman Gorge. The tripartite arrangement is overseen by the Family Responsibilities Board comprising senior representatives from both Governments and the CYI. Both governments have committed significant resources, with a combined investment of over \$100 million over four years.

The CYWR trial is groundbreaking, unique in the world, linking parental responsibility with government assistance. It represents a significant departure from previous government policies and has meant fundamentally reforming the way communities and governments operate to remove the disincentives that cause dependency cycles — which in turn has meant the Commonwealth and State cooperating centrally and on-the-ground in the communities to a previously unheard of level.

The benchmarking occurs through a quarterly report, provided for each community that includes data on:

-
- Magistrates Courts notifications
 - School Attendance notifications and school attendance more broadly
 - Child Safety notifications
 - Housing Tenancy notifications
 - number of conferences with the Commission
 - implementation of specific community programs and social services.

Comparisons can then be made over time, between communities, and against the broader State average. These quarterly reports are tabled in Parliament, but importantly they are being supplemented by an independent evaluation — which is commenting not only on the implementation of the model, but also on the effects of the model on individual and community well-being. This holds the potential of joint policy learning for the Commonwealth and State.

This is a long-term project. Mr Pearson's work is based on the notion of re-establishing social 'norms' and that's not something that can be achieved in a matter of months or even years. But in this instance, the benchmarking is being used as part of a bigger conversation.

As noted earlier, this is an uncommon situation. The success to date can be attributed to a number of factors:

- agreement that something had to be done and that 'business as usual' was no longer an option — including a recognition of the unintended consequences of previous policies
- a framework developed by CYI — that is, some good solid thinking to inform a new approach, that came from outside of government
- a public commitment by politicians, policy makers and service providers to improve the prospects for Indigenous children and families living in these areas, where indicators of social dysfunction, economic exclusion and wellbeing are among the worst in Australia
- tri-partite governance at a senior level
- a range of formal coordination agreements/arrangements including:
 - CYWR Program Office of senior officers from the three partners
 - Local Program Offices in each of the four communities
 - a Formal Partnership Agreement, signed in 2008 spelling out each partner's roles and responsibilities

-
- operating principle that all key decisions — funding, program development/delivery, recruitment, etc. must be agreed by three partners.

Such a situation may be difficult to replicate elsewhere.

13.6 Conclusion

This paper presents an unashamedly State perspective on benchmarking and its impact on Federal systems.

The recent COAG reform processes have challenged the States and Territories to focus, have a clear position (and as Queenslanders, we like to think of ourselves as having a unique position), and to act — to do worthwhile things that we might otherwise not have done, or not have done as quickly.

Benchmarking takes us from the rhetoric of reform to describing the actual changes on the ground that we believe will add up to better outcomes. It's valuable because it makes us think this through — articulate the concrete actions, outputs, and/or benefits that will improve life for the community — and then keeps us honest in our delivery by tracking performance.

However, as with all powerful mechanisms, benchmarking has the potential to be misused or to bring about unintended consequences. In and of itself, benchmarking is neutral; its impact is dependent on context and the way it is used. This paper contends that context is critical. Sensitivity to context has to be the mark of a good system — otherwise the opportunities for policy learning are lost.

Benchmarking takes place within the context of a relationship: it always comes back to relationships. When we get the relationships right we can achieve anything, but when they aren't tended to appropriately all sorts of problems ensue. And, of course relationships are never static: they require ongoing effort.

References

- 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.
- Forster, P. 2005, *Queensland Health Systems Review – Final Report*, The Consultancy Bureau, Brisbane.
- Schön, D.A. 1983, *The Reflective Practitioner: how professionals think in action*, Basic Books, New York.

14 Commonwealth–State benchmarking and the state of Australian Federalism*

Helen Silver¹

Victorian Department of Premier and Cabinet

14.1 Introduction

Like any good public servant, I want to start by managing expectations.

I think you would all agree that Commonwealth–State benchmarking and the state of Australian Federalism are not topics that naturally lend themselves to a high-wire comedy act. At the same time, however, you can rest assured that I will not try your patience with a basic seminar on benchmarking and federalism in Australia.

Instead, tonight I want to draw on my experiences from two of the defining Council of Australian Government (COAG) reforms of the past few years to make a more general, integrated and, I think, more interesting point. The two COAG reforms I will be drawing on are the post-2008 federal financial relations framework and the competition and regulatory reform agenda. In reflecting on these reforms and my related experiences of COAG meetings, there can be no doubt that these gatherings of government leaders in Australia have provided no shortage of personalities, fast-moving politics and grand drama.

We have had some vigorous debates on particular issues and we will have them again. That is the nature of robust public policy making and it should be welcomed. But beyond these immediate and attention-grabbing events, over the past years we have agreed on some fundamental backstage reforms to the way governments work together.

* Dinner Speech, Tuesday, 19 October 2010.

¹ Helen Silver is the Secretary, Victorian Department of Premier and Cabinet.

COAG has started to institutionalise the sort of policy and governance disciplines that we adopt within our own governments and that are also shaping best practice overseas. My main point tonight is that these shared practices — which I will describe as a strategic policy logic of outcomes-based, evidence-driven benchmarking and a genuinely federal approach to national governance — will survive the highs and lows of Commonwealth–State relations because, in short, good processes get good outcomes.

14.2 The common features and shared strategic policy logic of benchmarking and federalism

Before I draw on some COAG case studies to illustrate this central theme, I need to give credit where it is due. By bringing these two topics together, the Productivity Commission and the Forum of Federations have shown a lot of practical wisdom.

Without renegeing on my promise not to rehearse the benefits of benchmarking and federalism, it is useful to draw out their mutually reinforcing features and shared strategic purpose. We are all familiar with the rationale for benchmarking, in promoting public accountability, comparative learning and competitive performance assessment. Similarly the public benefits of a federal structure of government, in enabling flexibility, diversity, accountability, competition and innovation, are well known.

Clearly, benchmarking is not a passing fad, just as federalism is not an evolutionary stage on the road to unitary government or something that we can ‘fix’ once and for all. Instead, when we reflect on how benchmarking and federalism work together, we see that these approaches to public policy and national governance actually help us to understand and deal strategically with contemporary problems.

Taken together, benchmarking and federalism promote policy rigour, encourage good government and help us provide better outcomes for citizens. They also share important practical similarities, in that they both can be incredibly difficult in practice, their value is not always well understood by key stakeholders, and neither is done for its own sake.

My aim tonight is to explore these intersections and similarities and, in doing so, to demonstrate that a deliberate and integrated approach to benchmarking and federalism is simply part and parcel of being evidence-based in our policy analysis and self-consciously systematic in our governance arrangements.

14.3 How Vertical Fiscal Imbalance distorts Commonwealth–State relations and national strategic policy

Before I move on to my case studies, however, I need to name the ‘elephant in the room’: the excessive disparity that exists between the Commonwealth’s superior revenues and the States and Territories’ direct infrastructure, service delivery and associated spending responsibilities. We all know this elephant by its nickname, vertical fiscal imbalance (or VFI), and we all know that it is the main cause of difficult negotiations, blurred roles and responsibilities, and media and public misunderstanding.

For our international visitors tonight, let me briefly summarise the Australian version of this common federal fiscal dilemma. Some mismatch between a central government’s tax-base and regional governments’ spending responsibilities is not unusual, and might even be desirable. Unfortunately, in Australia the fiscal imbalance between the Commonwealth and the States and Territories is vast. The Australian federation has the dubious honour of competing for the most extreme VFI in the world. In Victoria nearly 50 per cent of our \$46 billion State budget comes from the Commonwealth. Some other Australian jurisdictions are even more dependent on the Commonwealth for revenue.

Excessive VFI has the potential to undermine an evidence-based and rigorous approach to the distribution of public accountability in a range of policy areas. It does not make centralisation inevitable, but it does encourage opportunistic appeals for federal interventions and can contribute to a principal–agent attitude to federal relations.

14.4 First case study: the IGA FFR

Having recognised these challenges, I’d like to turn to my first COAG case study: the 2008 federal funding reforms. This case study provides us with a valuable illustration of how — despite VFI — we have nonetheless started to institutionalise a better way of working together.

I’ll summarise these reforms briefly for our international visitors. The *Intergovernmental Agreement on Federal Financial Relations* — the FFR framework — has established a national outcomes-based funding and performance regime. It covers six key policy and services areas, including health and education, while also providing a clear articulation of the principles for future cooperation.

A mere two years ago, the Australian federation did not have a robust and reform-enabling framework for federal financial relations. Now we have a centralised process for administering payments to the States and agreed core funding on an *ongoing* basis. This means we should be able to avoid fights every five years over new funding agreements. The Framework also provides for reform pilot initiatives which, if successful, could subsequently be rolled into the core funding. These were hard won gains.

This focus on outcomes-based funding has been matched with greater performance reporting. The FFR framework empowers the independent COAG Reform Council (CRC) to publish performance information against outcomes annually for *all* jurisdictions. These reports are major steps towards better national benchmarking and more meaningful public accountability. Stakeholders, the national media and the general public are becoming more aware of them, and no jurisdiction — the Commonwealth or a State or Territory — will be able to hide when a CRC report reveals poor comparative performance.

The FFR framework continues the strategic policy logic of benchmarking and federalism, and it forces us to continually improve on the key outcome metrics that matter to the public. These reforms — the focus on outcomes rather than input controls, and the incentives for innovation — have been significant. But they have, at times, been lost in wider public debates on COAG and its reform agenda. Unfortunately, some reports in the national media present performance reporting under the FFR framework as an exercise in the ‘blame game by other means’.

Data quality issues in particular present a shared challenge, but often media reports on these issues are framed as the Commonwealth ‘pushing’ States and Territories to release information and raise their performance. In reality, working through these issues and refining agreements will take time and more pro-active governance by all jurisdictions.

We need, as the CRC Chairman has recently said in relation to their report on COAG’s overall progress, to sustain our efforts in fully implementing these far-reaching reforms. I am confident that the FFR framework will have a significant, long-term and positive impact on the quality of intergovernmental cooperation — and, in turn, a positive impact on the quality of Australian public policy and government services.

14.5 Second case study: the competition and regulatory reform agenda

For my second case study, I want to reflect on an important substantive component of COAG's agreed reform agenda — competition and regulatory reform. This case study takes our reflection on the shared strategic policy logic of benchmarking and federalism in a slightly different direction, by demonstrating how we need to apply these disciplines when weighing up — on a case-by-case basis — the relative costs and benefits of regulatory competition, mutual recognition, harmonisation and centralisation.

Prior to COAG's most recent reform agenda, it had been argued by some that competitive federalism in Australian regulatory systems had failed, and that States and Territories should instead resist parochialism and embrace market reforms in the national interest. Since then, COAG has made good progress on extending and completing its previous competition and business regulation reform agenda through a new *National Partnership Agreement to Deliver a Seamless National Economy*.

Reflecting on those debates and these cooperative reforms, it seems to me that arguments based on 'parochialism versus the national interest' do not do justice to either the case for national market reforms or the reform-enabling potential of competitive federalism.

As should be clear from what I have said tonight, our commitment to federalism is not based on a parochial or abstract commitment to States' rights. Instead, it is a commitment to context-sensitive, deliberative, accountable and right-sized governance.

Where a rigorous and evidence-based cost-benefit analysis supports a centralised approach — even a referral of powers — then of course that is the approach we should take. The national systems for business name registration and Standardised Business Reporting are good examples of this, just as the case for advancing a lot of the Seamless National Economy agenda was well established. More generally, and as scholars of federalism well know, the need for cooperation in a federal system can reinforce the case for such reforms, by demonstrating their broad-based support across governments and thus building their public legitimacy. COAG acted as a catalyst for change in this case and that is a positive message.

Equally, however, where the siren song of centralisation risks leading us into a uniform *but counter-productive* national regulatory regime, we should collectively pause and take stock of the real costs involved. There is nothing automatically more efficient about having uniform and centrally-controlled regulatory regimes for every product or service market that has a national dimension. A uniform regime, that adopts the wrong regulatory settings or approach, can be much more costly to the

national economy than having eight separate regimes. Equally, the choice to have a uniform national scheme necessarily stifles innovation, both by preventing jurisdictional experimentation and by potentially requiring a further round of multilateral negotiations before a cooperative scheme can be adjusted in light of experience or changing circumstances.

Without pre-judging the case, we should always start our shared thinking from first principles, by clearly articulating our common regulatory goals and weighing up the costs and benefits of how to get there. In this area, like others, we should apply good strategic policy logic by drawing on real-world evidence and being self-consciously systematic in our governance arrangements.

14.6 Concluding thoughts

In closing, I want to return to the practical wisdom in the theme for this roundtable. We have learned a lot about getting the most out of benchmarking and federalism in the last few years.

The COAG reforms that I have discussed tonight demonstrate the growing role in the Australian federation for a strategic policy logic focused on outcomes and facilitated by evidence-based benchmarking and a deliberative approach to national governance. Overall, I think we have started to embed the shared institutions and culture upon which the governments of our federation can develop and deliver better policy and service outcomes for all Australians.

Given my emphasis tonight on how good processes support good outcomes, it should not surprise you that I regard ongoing institutional reform of COAG itself as important. COAG provides the governments of Australia with a shared strategic decision-making and coordination forum. To deliver on this role, I would like to see COAG adopt and formalise some basic procedural disciplines, such as planning for a small number of regular meetings each year. Similarly, COAG needs an independent secretariat to coordinate a more focused agenda and allow for the States and Territories to put issues on the table for discussion and action. To underpin such changes, I think an intergovernmental agreement enshrining COAG's principles and governance would be a very positive step.

The undeniable merits of good processes in supporting good outcomes are such that I think systemic reform of COAG's operations, and a better focus of our collective efforts on key shared national challenges, is a real possibility in the near term.

I hope my thoughts this evening are useful to your conversations over the next two days. I look forward to hearing the outcomes of these productive discussions.

A Roundtable program

Day 1 – Tuesday 19 October 2010

16:00–16:30 **Welcome**
Gary Banks, Chairman, Productivity Commission
Felix Knüpling, Director, Europe Programs and Australia, Forum of Federations

Session 1 ***Federalism and benchmarking in comparative perspective***

16:30–18:00 **Prof. Alan Fenna**, Curtin University, Perth

Discussion

219:00–22:30 **Dinner**

Speaker: **Helen Silver**, Secretary, Department of Premier and Cabinet, Victoria on *Commonwealth–State Benchmarking and the State of Australian Federalism*

Day 2 – Wednesday 20 October 2010

16:00–16:30 **Welcome**
Gary Banks, Chairman, Productivity Commission
Felix Knüpling, Director, Europe Programs and Australia, Forum of Federations

Session 2 ***Benchmarking as an accountability tool***

The morning sessions, moderated by Felix Knüpling, focus on international case studies.

8:45–10:15 **Prof. Kenneth K. Wong**, Chair, Department of Education, Brown University

Dr. Clive Grace, Honorary Research Fellow, Centre for Local & Regional Government Research, Cardiff University, former Director General of the Audit Commission (Wales)

Discussion

10:15–10:30 **Coffee break**

Session 3 ***The open method of coordination of the European Union in action***

10:30–11:30 **Bart Vanhercke**, Co-Director of the European Social Observatory (OSE), former European advisor to the Belgian Minister of Social Affairs

Discussion

Session 4	<i>Emerging benchmarking exercises</i>
11:30–13:00	<p>John Wright, President and CEO of the Canadian Health Information Institute (CIHI); former Deputy Minister of the Government of Saskatchewan</p> <p>Dr. Gottfried Konzendorf, Federal Ministry of the Interior</p> <p>Discussion</p>
13:00–14:00	Lunch
Session 5	<i>The Australian experience with federalism and benchmarking: RoGS</i>
	The afternoon sessions, moderated by Gary Banks, focus on the Australian experience.
14:00–15:00	<p>Gary Banks, Chairman of the Productivity Commission</p> <p>Lawrence McDonald, Head of Secretariat for the Review of Government Service Provision, Productivity Commission</p> <p>Discussion</p>
Session 6	<i>Benchmarking and the Council of Australian Governments (COAG) Reform Council</i>
15:00–15:45	<p>Mary Ann O’Loughlin, Executive Councillor and Head of Secretariat, COAG Reform Council</p> <p>Discussion</p>
15:45–16:00	Coffee Break
Session 7	<i>Australian perspectives on benchmarking</i>
16:00–17:15	<p>Sharon Bailey, Executive Director, Office of the Director General, Department of Premier and Cabinet, Queensland</p> <p>Peter Dawkins, Secretary, Department of Education and Early Childhood, Victoria</p> <p>Ben Rimmer, Deputy Secretary, Strategic Policy and Implementation, Prime Minister and Cabinet</p> <p>Discussion</p>
17:15–17:45	<p>Wrap-up and closing remarks</p> <p>Prof. Alan Fenna, Curtin University, Perth</p> <p>Felix Knüpling, Director, Europe Programs and Australia, Forum of Federations</p> <p>Gary Banks, Chairman, Productivity Commission</p>

B Roundtable participants

Speakers

Felix Knüpling	Director, Europe Programs and Australia, Forum of Federations
Gary Banks	Chairman, Productivity Commission
Lawrence McDonald	Head of Secretariat for the Review of Government Service Provision, Productivity Commission
Helen Silver	Secretary, Department of Premier and Cabinet, Victoria
Peter Dawkins	Secretary, Department of Education and Early Childhood Development, Victoria
Sharon Bailey	Executive Director, Office of the Director General, Department of the Premier and Cabinet, Queensland
Ben Rimmer	Deputy Secretary, Strategic Policy and Implementation, Department of the Prime Minister and Cabinet
Mary Ann O'Loughlin	Executive Councilor and Head of Secretariat, COAG Reform Council
Alan Fenna	Curtin University, Perth
Kenneth K. Wong	Chair, Department of Education, Brown University
Clive Grace	Honorary Research Fellow, Centre for Local and Regional Government Research, Cardiff University
Bart Vanhercke	Co-Director, European Social Observatory
John Wright	President and CEO, Canadian Health Information Institute
Gottfried Konzendorf	Federal Ministry of the Interior, Germany

Participants

Lisa Gropp	First Assistant Commissioner, Principle Advisor Research, Productivity Commission
Scott Wheeler	Director, National Reform Branch, The Treasury, New South Wales
Sandy Pitcher	Acting Deputy Chief Executive, Cabinet and Policy Coordination, Department of the Premier and Cabinet, South Australia
Jenny Coccetti	COAG Performance Monitoring Unit, Department of the Chief Minister, Northern Territory
John McCormick	Director, Policy Division, Department of Premier and Cabinet, Tasmania
Trevor Sutton	Deputy Australian Statistician, Social Statistics Group, Australian Bureau of Statistics
Sara Glover	General Manager, Data Outcomes and Evaluation, Department of Education and Early Childhood Development, Victoria
Toby Robinson	Advisor, COAG Unit, Department of the Prime Minister and Cabinet