



**University of  
St Andrews**



## **Public Services Benchmarking under Localism and Federalism Sharing International Experiences**

**22<sup>nd</sup>-23<sup>rd</sup> February 2012**

**Local Government House, Smith Square, Westminster, London**

### **CONFERENCE SUMMARY**

#### **Introduction**

Benchmarking of local public services is critical for governments and communities who need to know whether services are effective and efficient and whether they are operating in the interests of their pupils, patients and clients. Benchmarking has become a central instrument for improving the performance of the public sector as under the right conditions, comparison can be an important driver of performance.

The UK in particular has been at the forefront of developments in this area. External performance assessment played a pivotal role in the Labour governments' strategy for public services reform. In contrast, the devolved administrations have developed more consensual approaches to assessment, such that the different methods adopted within the UK over the last ten years provided a 'natural laboratory' for comparing the effects of 'co-operative' and 'competitive' approaches. The 'localist' approach of the UK Coalition Government moves us into uncharted waters. It assumes that local authorities will scrutinise their own performance and voters will make 'rational' choices when presented with performance data.

Benchmarking and external performance assessment is a transnational phenomenon and there is much to learn in this new environment through knowledge exchange with other countries, and particularly federal jurisdictions, where there is keen interest in comparative public services performance. For example, Australia has recently adopted benchmarking as part of a major redesign and cut back of conditional grant programmes and Canada has been working on an integrated health information system.

While the basic idea of benchmarking is a simple one, it can be complex and challenging to put it into practice. In this context, this conference aimed to exchange knowledge and experiences between policy makers, practitioners and academics from Australia, Canada, and across Europe. The conference was part of a wider project co-funded by the ESRC and the Forum of Federations.

The conference set out to address a series of questions including:

- How is the impact of performance assessment currently assessed and how can it be measured with greater accuracy and confidence?
- What are the best ways for national and local governments to stimulate and support improvement in local services through performance assessment?
- What are the best international practices, what lessons can be learnt from them, and how can experiences be translated into the UK context?
- How do theories of improvement vary across different contexts and are these reflected in performance assessments?
- How do methods of assessment need to be adapted to optimise their effectiveness in different localities and service settings, and how do assessment methods best connect to other instruments of public service improvement which states routinely deploy?
- How can methods of performance assessment be best tuned to the needs of citizens-as-assessors, and also to key issues such as public expenditure reductions and the need for innovation?

### **The International and the UK Perspectives and Context**

**Professor Tony Travers** from the London School of Economics, UK chaired the opening session. He outlined the importance of benchmarking and explained that it is well developed in different places. It is particularly important at a time of a global recession to bring countries together to learn from each others' practices in delivering efficient public services. He noted that benchmarking is a helpful tool in that respect but that it is very complicated to make comparisons within countries and even harder to do so from country to country. He concluded by recognising the importance of exchange programmes, especially between the UK and federal countries, in order to understand different complexities and learn from different experiences.

**Professor Alan Fenna** from Curtin University, Australia, is the lead expert on the Forum of Federations' Benchmarking in Federal Systems programme. He has been studying how benchmarking has been developing and interacting in different forms of governance in federal systems. He explained that federalism and benchmarking are closely related. Benchmarking is a key device for managing inter-governmental relations. Federalism provides an opportunity for policy learning as sub-national units can pursue their own policies and other units can learn from their experiences.

Prof. Fenna outlined two main models in which benchmarking have been practiced. The first occurs when there is a benchmarking entity above a government. For example, the OECD has set up a variety of benchmarking studies across its member states but as it does not exercise any authority among its member states, there are concerns about the impact of this sort of benchmarking on public policy (although see the recent exception of PISA scores). The second model is where a unitary state imposes benchmarking on its' subordinate states such as in the UK. In between those two extremes lies federalism, defined as a system where the

nation state is made up of a national government and constituent units' with their own semi-sovereign status and autonomous policy jurisdiction; therefore they cannot have policies imposed by sanctions.

Prof. Fenna said that there are differences in benchmarking at the national and sub-national level, as well as a variety of obstacles to applying it in the public sector. One such obstacle is the tension between measuring efficiency of outputs in a context when we really care about is outcomes. He reported a healthy scepticism on benchmarking systems which may 'work' but are of little use if they don't lead to improved outcomes.

## **Benchmarking Local Public Services in the UK**

**Professor Steve Martin** from Cardiff Business School, UK explained that the UK has a history of benchmarking of local government performance based on a range of techniques including statutory performance indicators, external inspections, excellence schemes and peer review.

The Blair/Brown Governments had sought to drive service improvement using performance assessment frameworks including the Best Value regime, Comprehensive Performance Assessments and Comprehensive Area Assessments, all of which were based loosely on balanced scorecard approaches. Drawing on research conducted with James Downe, Clive Grace and Sandra Nutley, Prof. Martin explained that these frameworks could be conceptualised as performance improvement regimes that were shaped by interactions between central government, local authorities and audit bodies which consisted of a range of control instruments and reflected explicit theories of improvement.

These concepts could be used to map approaches to performance benchmarking and analyse similarities and differences – over time and between countries. Comparisons of the local government performance improvements regimes operating in England, Scotland and Wales had shown that they had common policy goals and used similar control instruments, but that differences in the relationships between the key institutions produced differences in the way in which regimes were designed and implemented. The success of benchmarking could be measured in terms of process, programmatic and political outcomes. This was an underdeveloped area of research in which exchanges of experiences between federal and unitary systems would be particularly interesting.

## **Benchmarking in Education and the Environment**

### ***Education***

**Professor Kenneth Wong** from Brown University, USA discussed the promise and challenge of benchmarking in Education. He started by highlighting the fact that there is nothing in the US constitution on education – it is a federal responsibility. According to Prof. Wong, there have been three phases of federalism. The first phase saw the growth of categorical grants where the federal government provided the resources for programmes which were delivered by the states. In phase two, there was a growing emphasis on performance and accountability. With the No Child Left Behind Act (NCLB) of 2001, the federal government aimed at combining an input-based framework with outcome-based

accountability to assess whether children were meeting the standards. The federal government set a goal of 100% proficiency by 2014 and it was up to individual states how they reach this target. Almost inevitably, states have made uneven progress, so there are sanctions in place which intensify over time if standards are not met. These sanctions vary from writing a letter to parents advising them that better performing schools are available, to principals being fired or the potential for a whole district to be taken over by the federal government.

The third phase of federalism is where accountability mechanisms have been supplemented by institutional innovation. Prof. Wong explained that the American Recovery and Reinvestment Act (ARRA) (2009) marked a historic shift as it doubled the federal contribution of total spending in public schools as well as creating the policy conditions for innovative reform and stronger accountability on student performance. President Obama introduced 'Race to the Top' where states are invited to submit their best ideas on system transformation and school innovation. The early winners of this competition focused on the value of improving teacher accountability as a way to transform public education.

Prof. Wong concluded that there has been an era of unprecedented federal activism. Schools can no longer hide behind data as there is more transparency and public reporting. There is greater local/state support for charter schools, a reliance on state leaders as change agents and local reluctance in school turnaround and performance-based compensation reform.

**Dr Deborah Wilson** from Bristol University, UK discussed the assessment of school performance in England and Wales. She explained that the performance assessment of schools started in 1988 (Education Reform Act) and that there is lots of publically available data to show how schools are performing. The media have been creating school league tables for more than two decades. There is a large body of evidence on the ways in which schools respond to this agenda which include gaming. There has, therefore, been some change in performance indicators (PIs) which have included 'Value added' to reduce incentive for 'cream-skimming' and including English and Maths in the percentage of students achieving 5A-C grades to mitigate the choice of alternative qualifications. There is less evidence on the effect of league tables on school effectiveness.

Dr Wilson's presentation focused on examining the performance of schools in England which has a competitive system versus a more consensual system in Wales where there are no league tables. The hypothesis is that if performance changes (and everything else is kept equal), any change is due to the publication of league tables. Her results showed that school effectiveness fell in Wales relative to England after league tables were abolished by almost two GCSE grades per student per year. The impact was concentrated in the lower 75% of schools so that average performance fell and educational inequality rose in Wales relative to England after the reform. The results are mirrored by PISA results for England and Wales. The Welsh Government did not welcome the results but have recently introduced a banding system for schools so there is opportunity to produce league tables.

Dr Wilson concluded that performance indicators provide opportunities for undesirable as well as desirable responses and that encouraging explicit comparison of performance provides incentive for both. While the role of PIs in enabling choice is often emphasized, there is little evidence to suggest that choice improves outcomes.

## *The Environment*

**Daniel Wachter** from the Federal Office of Spatial Development, Switzerland made a presentation on the performance of cantons and cities on sustainable development. There are 26 cantons in Switzerland which vary significantly in size from 50,000 to 1.2 million people. The Cantons are sovereign except that their sovereignty is limited by the federal constitution. They shall exercise all rights that are not vested in the Confederation. The system has, therefore, been described as the most decentralised in the world.

There is a specific article in the constitution on sustainable development. A federal government report suggested that 'Cantons and municipalities are to receive support with the development and deployment of suitable instrument for monitoring, managing and evaluating sustainable development'. As a result, the 'Cercle indicateurs' is a forum dedicated to the development and use of sustainability indicators for Swiss cities and cantons. This involves 19 cantons and 16 cities and includes a common set of core indicators (n=37) to compare levels of performance. Cantons and cities provide the performance data and some commentary on the results and the Federal Statistical Office publishes them on their website. The analysis allows cantons and cities to compare themselves to others and examine their relative strengths and weaknesses. There is no legal obligation to be involved and seven cantons do not currently participate. It is a form of collegial benchmarking as the indicators are designed locally and there are no sanctions or incentives.

There are a number of lessons from this benchmarking exercise. First, collegial approaches can be appropriate where central government does not need to or cannot exert control – compare this process to Prof. Wong's presentation above. Second, you need to have clearly defined objectives, rules and governance in place for every benchmarking project. Third, methodological and technical quality is crucial but there also needs to be some flexibility in response to data availability. Finally, having a participatory approach can secure long-term support for the process.

**Professor Tim Allen** from Birkbeck University, UK examined local performance management and benchmarking in English local environmental policies and services. He started with the premise that many established principles and practices in benchmarking and performance management are applicable to environmental services and activities as they are to other local public services. He explained that English environment policy responses primarily emanate from national and international policy and legislation, with local responsibilities delivered through a mix of central government owned agencies, local government and voluntary sector organisations.

Environmental benchmarking has primarily been a response to this centrally driven agenda, with a focus either on managerial and cost measures of efficiency and effectiveness, or on the application of national indicators to localities (such as National Indicator 186: 'Per capita reduction in CO2 emissions in the local authority area'). These measures have a role but do not reflect the complexities of assessing performance in local environmental policy and services. One problem is that benchmarking practice lags and there is no clear consensus on what to benchmark or how to respond to the need for solutions that dictate multi-agency engagement and action. Another issue is indicators responding to legislative pressures such as the emphasis on increasing recycling when more emphasis could be placed on reducing product packaging.

Prof. Allen concluded that benchmarking has the potential to improve environmental performance management as it can assess effectiveness and productivity against well defined service or performance objectives. It can be a means to compare measurable indicators of performance to demonstrate progress against well-defined goals such as energy consumption and therefore be a diagnostic tool to trigger inquiry. To be effective, however, benchmarking requires open collaboration and co-operation between organisations and sectors and meaningful data that is robust and transparent.

## **Benchmarking in Health and Social Care**

**Dr Jeremy Veillard** from the Canadian Institute of Health Information discussed the experience of benchmarking in health care in Canada. There are in effect 14 different systems of health care in Canada: ten provincial systems, three territorial systems and a federal system. The federal health system is limited to few responsibilities and provinces have jurisdiction over almost all hospitals and exclusive responsibility for the delivery of most medical services. In terms of data collection, the federal government holds the national health accounts and supporting data. In 1994, the Canadian Institute for Health Information (CIHI) was created as an independent organisation through funding and support from all jurisdictions. It has a mandate to consolidate provincial health data and it is the main source of pan-Canadian indicators of performance.

There are ongoing concerns about the affordability of Canadian health care, 70% of which is funded from the public purse. Some 80% of public funding is provincial and 20% federal. Federal funding transfers are nominally aimed as cost sharing and the maintenance of equity across the country. These transfers are an ongoing source of tension between federal government and the provinces. However, a series of First Minister Health Accords have resulted in increased federal funding in exchange for greater performance transparency and public reporting.

Despite some early resistance to implementation, provinces have developed comparable indicators, and there is general interest in performance measurement, accountability, benchmarking and capacity building. There are ongoing challenges relating to data collection and quality, the timeliness of reports, and understanding the 'best practices' underpinning good performance. New areas of development include the emergence of structured benchmarking activities with a policy focus and initiatives to improve the timeliness, accessibility and understandability of performance data.

**Bart Vanhercke** from the European Social Observatory explained the process of benchmarking social care in Europe through the Open Method of Coordination (OMC). The OMC is a relatively new regulatory instrument introduced by the Lisbon Council in 2000. It is a mechanism for coordinating domestic policies in areas where the EU has no formal competencies, such as in social protection and inclusion, pensions, health and long term care. The OMC is based on agreeing and establishing common objectives, key issues, performance indicators, stakeholder networks, peer reviews, and joint reports, which include 'recommendations' to member states. A variety of indicators have been developed which enable the measurement of policy challenges, success and failures in social care.

Vanhercke argued that the social care OMC has had an impact on domestic and EU policies through leverage and policy learning. Performance indicators are being harmonized to make

them comparable across member states and the creation of a multidimensional indicator set has been important. The development of two different types of peer review within the OMC is considered to be an essential learning mechanism, but there are concerns that these reviews should be more closely linked to policy reforms.

**Professor Gwyn Bevan** from the London School of Economics, UK discussed models of governance and benchmarking in health care. He outlined four models of governance and discussed their potential strengths and limitations in principle. ‘Altruism’ relies on individual motivation and it is known to be both popular with citizens and a low cost approach to improving public services. However, it can generate perverse initiatives by inadvertently rewarding failure. ‘Hierarchy and Targets’ is often characterised as powerful in the short run, however it can be a victim of gaming behaviour, which can make it dysfunctional in the long run. ‘Choice and Competition’ is often assumed by proponents to improve the responsiveness and performance of public services, but there are problems of design and implementation which inhibit this potential. The ‘Reputation’ model focuses on naming and shaming through the production of published rankings. In this model, transparency of performance is important but this may be difficult to achieve because the approach is also susceptible to gaming.

Drawing on evidence from the USA, UK and Italy, Bevan argued that a reputation approach to governing health care has been more effective in improving performance than the other models. For example, he used 2000-2005 UK data to compare and contrast the performance of England, which published star rankings for health care organisations in this period, with the performance of the devolved nations (Scotland, Wales and Northern Ireland), which mainly relied on an altruism model of performance improvement. He found that England consistently outperformed the devolved nations. Despite such evidence, reputation frameworks are not popular and the current UK government has moved away from this approach and is focusing more on a choice and competition model. Bevan argued that this is not a good move because the choice and competition model is also unpopular and is not likely to be as effective as the reputation model.

**Dr José Luis Fernández** from the London School of Economics, UK discussed the case of benchmarking in social care in the UK. Here social care includes the provision of social work, personal care, protection or social support services to children or adults in need or at risk, or adults with needs arising from illness, disability, old age or poverty and their families and carers. Fernández explained that in many cases, the aim of social care is to maintain wellbeing in the context of natural deterioration. Thus, social care can be characterised as an ‘experience good’ where the experience of those receiving care is often more meaningful than final outcome data. The majority of social care services in the UK are provided by the private sector. In this situation, performance assessment is important because of possible market failures, the vulnerability of individuals consuming services, and the benefits of highlighting and sharing good practices across the system.

Current social care benchmarking frameworks often seek to combine three objectives or elements: a punitive element (punishments for poor performance); an incentivising element (encouraging performance improvement); and an information element (providing information to stakeholders). There is increasing disenchantment with top-down, target-driven performance management and a desire to increase the focus on final outcome measures, but this is a challenge because raw outcomes data tend to provide a distorted view of

performance. On the basis that social care is often an experience good, there is a move to implement survey-based performance assessment but this is not straightforward.

Overall, Fernández concluded that there are many challenges to social care benchmarking. There are the familiar challenges of improving the quality of performance indicators, generating better methods for collecting data, and dealing with the risk of gaming. There are also challenges relating to the personalisation agenda in social care, where individuals have budgets to purchase their own care packages. This poses the risk of developing a two-tier system, with more stringent performance assessment for formal care providers, but almost nothing for individual providers/personal assistants.

## **Performance assessment frameworks in the UK and in international jurisdictions**

**Professor Colin Talbot** from Manchester University, UK explained that the explosion of the use of performance frameworks occurred in the mid-1980s where the public sector adapted and adopted a variety of performance frameworks from the private sector. The most prominent models adapted to the public sector were the European Foundation for Quality Management (EFQM) and the Balanced Scorecard (BSC).

Prof. Talbot stressed that over time new models have been developed in the public sector at different levels of government. He argued that on the one hand, some models were developed and imposed by central government and/or audit inspection or quality bodies, for instance, the 'Logic Model' or CPA in English local government. On the other hand, some models were innovated by individual public agencies, or sometimes groups of agencies, such as the South Wales performance framework for social services.

Prof. Talbot argued that a variety of performance frameworks have been created and implemented in different ways around the world, depending on various socio-economic, cultural and political realities. For instance, while in the UK the Audit Commission was in charge of collecting information and producing certain performance indicators (top-down process), local authorities in Poland developed their own data collection without any central government intervention. Finally, Prof. Talbot argued that most of the performance frameworks are created by circumstantial process. Few frameworks come with an explicit theoretical underpinning or any empirical testing or data to support them. He suggested that performance frameworks should be created by public policy specialists, practitioners and politicians.

**Lawrence McDonald** from the Productivity Commission, Australia presented the case of Reports on Government Services (RoGS) in Australia. He gave an overview of the Australian federation system which comprises six states, two territories and local governments. The states are sovereign entities, subject to certain powers of the Commonwealth as defined by the constitution. There is a strong centralising trend which has been driven by vertical fiscal imbalance. The commonwealth provides about 50% of the states' and territories' revenues. Performance assessment has improved through Reports on Government Services and National Agreement reports at the state level and the Victorian Auditor General work at the local government level.

McDonald acknowledged that performance assessment has improved services. The process where jurisdictions learn about their own and other jurisdictions has clarified objectives and



generated an 'evaluation' of inputs. Agencies have improved in the management of information and the validation of performance, and the public has benefitted from the publication of performance information and increased accountability. One of the main benefits of the RoGS is the measurement of the impact and benefits of national reforms, which supports the analysis of efficiency and effectiveness of public policies. In terms of performance management in local governments in Australia, McDonald highlighted the Victorian State experience where the Auditor General is currently conducting an inquiry into performance reporting by Victorian local governments.

**Rabea Hathaway** from the Federal Ministry of the Interior, Germany introduced their benchmarking experience. Germany is composed of the Federal government and 16 states (Länder), which are partly sovereign constituent states. The states are represented at the federal level through the Bundesrat, which is the upper house of the German parliament. Unlike other federal countries, Germany has been developing a form of 'cooperative federalism' where decisions are made by various decision-making levels working together.

The constitutional reform of 2009 gave benchmarking a constitutional status. It calls on federal and state administrations to conduct comparative studies. In practice, however, the implementation has been handled differently by the states. At the state level, they are responsible for administrative organisation and drafting a plan for conducting comparative studies in public administration. But 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. In many cases, benchmarking has been rejected due to lack of comparable basic data.

Germany has some experience of benchmarking at the local government level where 7 to 10 municipalities had the common aim of determining, comparing and discussing conclusions about public services data. Comparative studies groups at local level are well known for their experience and coordination. Some areas of comparison are personnel management or public service offices.

Overall, all interested parties in Germany agree on the importance of benchmarking and the necessity of anchoring it in law and institutions. However, the actual implementation on federal and state level is proving to be much more difficult than was envisaged by those who wrote the article binding benchmarking into the German constitution.

**Mark McAteer** from the Improvement Service (IS), Scotland introduced the local government performance framework experience in Scotland. The IS was set up in 2004 and is a partnership between the Convention of Scottish Local Authorities (COSLA) and the Society of Local Authority Chief Executives (SOLACE). It is a company limited by guarantee funded through local government settlement and its' members are the 32 Scottish councils. Its purpose is to help improve efficiency, quality and accountability of local public services in Scotland by providing advice, consultancy and programme support to councils and their partners. The majority of councils' funding is from the Scottish government and through the Council Tax that the council itself sets.

The benchmarking project of Scottish local government aims to drive improvement through a comparative performance support framework and to support councils in targeting transformational change resources to areas of greatest impact (both of efficiency (costs) and productivity (outcomes)). The project is focused on the largest spending services plus

corporate services. It was initiated in 2010 and developed service grouping for managing the analysis, and identifying indicators comparable across councils and services which materially help to improve performance. Measurement was focused on both inputs and outputs. Indicators focus on where data is available, reliable and consistent. In terms of outputs, agreed indicators were created from useable data sources, divided into a set of family groups (socio economic, population, education, and so on). The IS has been improving data collection and standardisation, but there is a need to improve outcome data. A Benchmarking guide (toolkit) has been created in order to inform people what each benchmark is and how they can compare performance in the local authorities. The support of national government is important too, not only in terms of data collection but also in terms of political will for performance improvement.

McAteer suggested that it is very important to publish benchmark data on local government improvement and on failure. But given the variety of Scottish councils, there should not be standardised public performance reporting across local governments to avoid simplistic league tables. Local government needs to be open to mutual exchange and learning in order to improve services as well as to encourage confidence building and getting systems and culture right. There is a need to strengthen analytical capacity and to encourage benchmarking across partnerships, linking performance to outcomes and challenging service delivery options. This needs to take place with the active engagement of customers and politicians.

### **Data/information validity, verification and consistency**

**Professor Steven Van de Walle** from Erasmus University, Rotterdam examined the subjective assessment of performance. Public sector organisations and academics have been adapting and adopting performance frameworks from the private sector to the public sector in order to assess inputs, outputs and outcomes of a variety of public services but what is being measured? Benchmarking has tended to focus on inputs and outcome indicators but there is only limited robust outcome data available. Government and public sector organisations have used surveys as proxies of outcomes but this raises important questions. How was the data collected? Who was asked and why? How relevant is the data? Does this outcome data reflect how public services are performing? The answers to all of these questions are important in determining the usefulness of that data.

Prof. Van de Walle showed that there is a negativity bias in responses with exceptional events receiving more attention than common ones. The way in which you frame questions may also influence the results. For example, a communication survey in Canada on a general question, 'How would you rate the performance of the government of Canada?' resulted in 29% of the sample answering that it was 'good'. When the question was reframed as 'Now that you have had an opportunity to think about the Government of Canada's priorities and performance in more detail, I would like to ask you again - generally speaking, how would you rate the performance of the Government of Canada?', the response was 40% replying 'good'.

**Andrew Stephens** from the Local Government Data Unit, Wales introduced the Welsh local government performance improvement framework which aims to increase the quality and availability of data used to support improvement. Stephens explained that data collection is based on a range of measures and that data quality in benchmarking is vital to support

improvement and public accountability. Performance indicators have to be relevant to the services and driven by central collection and validation.

The Benchmarking model implemented in Wales begins with a review of data from specialists on data issues, followed by an examination of the real differences in service performance, which leads to agreements of actions towards service improvement which are reviewed on an ongoing basis.

**David Walker** from the Economic and Social Research Council (ESRC), UK stressed the importance of accurate and reliable data in the process of performance assessment and benchmarking in public services. He recognises, however, that benchmarking faces a variety of challenges such as complexity cost; accountability trades off against performance management; data needs subordinated to spending control; and trades off against data quality. One of the main drawbacks to publishing performance data is the absence of data literacy, not only by public sector organisations, but also by citizens. Benchmarking is inherently quantitative, therefore there is a need to understand data in order to analyse, compare and contrast performance. Public sector organisations should be aware of this potential problem and produce information in such a way which allows for citizens to easily compare local performance and spend data, but also to be useful for organisations to compare performance internally and against other organisations.

Walker highlighted the importance of an experienced and independent audit body. In the UK, ‘armchair auditors’ have the opportunity to examine local government performance, but experienced and qualified audit units are needed to evaluate issues such as productivity and efficiency. Benchmarking is therefore an uneasy bedfellow for localism as public capacity is inherently limited. Benchmarking needs an external authority to audit its performance.

**Professor Michael Hughes** from the Audit Commission, UK introduced his observations on validity, verification and consistency of data. Over two decades, the Audit Commission has collected, cleaned, analysed and published data. The Commission recognised the importance of different uses of data such as measuring progress, promoting accountability, and identifying opportunities for improvement. National, local and management indicators are beneficial not only for the public but also for the government, politicians, senior managers and stakeholders.

Prof. Hughes suggested that the Commission experienced some concerns about the quality of information and lack of knowledge in local government of how to create data, analyse and present information. This led to the Commission setting six standards for data quality - accurate, valid/free of manipulation, reliable, timely, relevant and complete – and concluding that there needs to be a commitment to a data quality culture, clear and well-disseminated policies to secure good quality data to support decision-making and accountability, a professional process, and data literate staff.

Prof. Hughes concluded that performance information should be relevant to the decision in hand, supported by data quality standards and be presented in a way which aids understanding.

## Conclusions

Benchmarking is a practical tool aimed at improving performance. Over the last few years, a variety of performance frameworks and benchmarking systems have been implemented among federal and unitary governments as part of a wider strategy for improving public sector performance. At this conference, Vanhercke explained that the Open Method of Coordination which uses multidimensional indicators has been able to compare and encourage improvement between the European member states. In Australia, McDonald highlighted the work of the Productivity Commission in producing Reports on Governments Services which provide benefits of comparative performance in a variety of social services. In Canada, the government has strong intergovernmental relations between its health authorities where benchmarking has been negotiated bilaterally and provinces report to a third party. While one might expect benchmarking in federal government systems to be very different to benchmarking in unitary states, the conference revealed a lot of commonality in the questions being asked, the problems faced, and the approaches adopted. There is much that we can learn about inter-governmental benchmarking by looking across federal and unitary systems. Little is known about how different benchmarking systems are developed, their theory of improvement, their use as an instrument of governance (particularly in federal systems) or how they fit with other improvement mechanisms. An area for future research would be to explore whether the different systems were the result of national, political, financial and social concerns and/or a lack of policy learning. Research could also focus on characterising the main differences between benchmarking designs by analysing them along a series of dimensions. Prof. Fenna's presentation offered one such dimension - a continuum from top down/coercive benchmarking to bottom up/consensus benchmarking. Another dimension to emerge from the presentations is peer-review versus indicator-driven approaches. It should be recognised, however, that this dimensional analysis would only take us so far because as Prof. Martin outlined, systems with similar dimensional profiles (CPA, BVA and WPI) can *feel* very different.

There is a particular need for research which examines empirically the impact of different performance frameworks and benchmarking systems as the existing evidence base is weak. The evaluation challenge is considerable because we need to develop a better understanding of what works, for what purposes, how, when and why, and with what spillovers and opportunity costs. At the moment, the limited evidence we have from both the education and health and social care sectors (Prof. Bevan and Dr's Wilson and Fernández) suggests that benchmarking systems that are based on hierarchy, targets and reputation effects ('naming and shaming') have the most impact on performance improvement, but they can be politically unpopular and hence are likely to have limited lives.

All benchmarking systems face considerable challenges in creating and capturing robust and comparable indicator data. There are a number of important points to bear in mind in trying to manage these challenges. First, robustness can only be judged in relation to the purpose for which the data is to be used (fitness for purpose). Second, it is difficult to standardise models of data collection because of the variety of interests (e.g. government and citizens). Third, most benchmarking systems capture historical rather than real-time data. Fourth, benchmarking requires open collaboration and co-operation between organisations and sectors. Finally, producing good comparative data is only one step in the benchmarking process; an equally important step is to provide analysis and interpretation of this data as this narrative is likely to influence the impact of benchmarking.

Given the challenges outlined above, there has been a lot of scepticism about the operation and impact of various benchmarking systems. This focuses on the difficulty of capturing service (or organisational) quality and performance through a limited set of indicators, the opportunity for data tampering and dealing with the risk of gaming. Despite this, since the 1980s, in the UK and in several other countries some form of benchmarking of public services and government activity has seemed inevitable. Maybe now is the time to question that inevitability and carefully consider any unintended side effects of benchmarking. David Walker raised the question of whether benchmarking is the right ship for the cold and choppy waters engulfing the public sector in many countries.

Finally, what is the role of citizens and service users in benchmarking? Local government is the level of government closest to citizens and that provides an opportunity for engaging citizens in the process. The UK's Coalition Government has provided some rhetoric surrounding the localism agenda in England which makes assumptions about 'armchair auditors' but evidence to date suggests that, in practice, citizens and service users are often only marginal participants in many benchmarking systems. They may be surveyed for their views on service quality and performance, which, as Prof. Van de Walle highlighted can result in some highly questionable data. The public are rarely involved in discussions about what the indicators should be, what they mean, or what should change due to the results of benchmarking. There may be scope for incorporating softer forms of intelligence about service quality and performance into benchmarking systems by using social media (e.g. Dr. Fernandez mentioned a 'trip advisor' for social care), but there is a question mark about whether the public are sufficiently interested in getting involved and using the data.

Our future knowledge exchange activities, which include international conferences in Canada and Australia, seminars across the UK with policy makers, practitioners and academics and dissemination workshops, will discuss the questions raised in this conference during 2012-2013. We will report back on these discussions and engage with participants on the findings using the LGA's Knowledge Hub.

## **Participants**

### **Research team**

Dr James Downe, Cardiff Business School  
Prof. Alan Fenna, Curtin University, Australia  
Dr Clive Grace, Cardiff Business School  
Felix Knüpling, Forum of Federations, Canada  
Prof. Steve Martin, Cardiff Business School  
Prof. Sandra Nutley, University of St Andrews  
Ofelia Saavedra

### **Participants**

Prof. Tim Allen, Birkbeck, University of London  
Prof. Gwyn Bevan, London School of Economics  
Steve Blake, Her Majesty's Inspectorate of Constabulary  
Eddie Bradley, Northern Ireland Audit Office  
Dette Cowden, Scottish Government  
Nick Easton, Local Government Association  
Dr José-Luis Fernández, London School of Economics  
Angela Hands, National Audit Office  
Rabea Hathaway, Federal Ministry of the Interior, Germany  
Jane Holownia, Wales Audit Office  
Prof. Michael Hughes, Audit Commission  
Shehla Husain, Department for Communities and Local Government  
Reg Kilpatrick, Welsh Government  
Ben Lucas, 2020 Public Services Trust  
Graeme MacDonald, SOLACE  
Mark McAteer, The Improvement Service, Scotland  
Lawrence McDonald, Productivity Commission, Australia  
Lesley Smith, Department for Communities and Local Government  
Andrew Stephens, Local Government Data Unit, Wales  
Prof. Colin Talbot, Manchester Business School  
Prof. Wendy Thomson, McGill University, Canada  
Prof. Tony Travers, London School of Economics  
Prof. Steven Van de Walle, Erasmus University, Rotterdam  
Bart Vanhercke, European Social Observatory, Belgium  
Jeremy Veillard, Canadian Institute of Health Information  
Daniel Wachter, Federal Office of Spatial Development, Switzerland  
David Walker, ESRC  
Juliet Whitworth, Local Government Association  
Dr Deborah Wilson, Bristol University  
Prof. Kenneth Wong, Brown University, USA

# **Public Services Benchmarking under Localism and Federalism Sharing International Experiences**

**22<sup>nd</sup>-23<sup>rd</sup> February 2012**

**Local Government House, Smith Square, Westminster, London**

## **Conference Agenda**

**22<sup>nd</sup> February**

**12.30 – 13.30 Registration, Lunch/Coffee and Networking**

**13.30 – 15:00 The International and the UK Perspectives and Context:** Chair: Professor Tony Travers (LSE)

Professor Alan Fenna and Professor Steve Martin and Dr Clive Grace

**15:00 – 15.30 Coffee**

**15.30 – 17:30 Concurrent Sessions**

**Education and Environment:** Chair: Felix Knuepling

Professor Kenneth Wong (Brown University) - No Child Left Behind  
Daniel Wachter (Federal Office of Spatial Development, Switzerland) - Sustainable Development of Cantons and Cities in Comparison  
Dr Deborah Wilson (Bristol University) - Assessment of school performance  
Professor Tim Allen (Birkbeck University) - Environment

**Health and Social Care:** Chair: Professor Sandra Nutley

Jeremy Veillard (Vice-President, Canadian Institute of Health Information) - Benchmarking Health Care in Canada  
Bart Vanhercke (European Social Observatory, Brussels) – Benchmarking Social Europe through the Open Method of Coordination (OMC)  
Professor Gwyn Bevan (LSE) – Performance in Health Services  
Dr José-Luis Fernández (LSE) – Benchmarking in Social Care

**18:15 – 19.15 Drinks**

**19.15 – 21.30 Dinner**

**23<sup>rd</sup> February**

**9:00 – 10:45      Performance assessment frameworks in the UK and in international jurisdictions:** Chair: Professor Steve Martin

Professor Colin Talbot (Manchester University) - Performance Frameworks in the UK  
Lawrence McDonald, (Assistant Commissioner, Productivity Commission) - Reports on Government Services, Australia  
Rabea Hathaway (Deputy Head, Performance and Joint Service Delivery Division, Federal Ministry of the Interior, Germany) - A Benchmarking Approach in the Making  
Mark McAteer (Improvement Service Scotland) - The Local Government Performance Framework in Scotland

**10:45 – 11:00      Coffee**

**11.00 – 12.30      Data/information validity, verification and consistency:**  
Chair: Dr James Downe

Professor Steven Van de Walle (Erasmus University Rotterdam)  
Andrew Stephens Local Government Data Unit Wales  
David Walker, Chair Infrastructure Committee of the ESRC  
Professor Michael Hughes, Director of Research, Audit Commission

**12:30 – 13.00      Thematic Concluding Session and Round Up:** Chair: Dr Clive Grace

Dr James Downe and members of the Research Team

**13.00 – 13:45 Lunch**

**2pm – 3pm      Conference Postscript** – a demonstration of the Local Government Association's 'INFORM' web based tool for comparing local public services performance.