Perversity in public service performance measurement

M. Pidd
Department of Management Science, Lancaster University, Lancaster, UK

Abstract

Purpose – To discuss some of the reasons why performance measurement systems in public services can lead to dysfunctional consequences even when people operate with the best of intentions.

Design/methodology/approach – The paper draws together literature from the UK public sector, from writers in performance measurement and from cultural insights in anthropology to understand why some of these perverse effects occur.

Findings – Though many reasons are cited for public service performance measurement regimes, it is clear that control aspects dominate the others. This, when allied to an unthinking use of cybernetic metaphors, is what can lead to dysfunctionality.

Originality/value – The paper should appeal to those who wish to improve the performance of performance measurement systems in public services and to those who wish to understand why things can go wrong.

Keywords Performance measures, Performance management, Public sector organizations, United Kingdom

Paper type Research paper

Introduction

There can be little doubt that performance measurement is at the heart of the UK government's work to improve the delivery of public services. UK public services now work within a performance framework that aims to involve government ministers, senior civil servants and the staff who actually provide health care, education and other services. There are claims and counter claims about the effectiveness of the system that is in place, but there can be no doubt that it is in place. The UK Parliament's Public Administration Select Committee reported on this in its fifth report, entitled "On target? Government by measurement" (Stationery Office, 2003). In the health sector this performance regime is almost taken for granted, though there are many suggestions (Bevan and Hood, 2004) about how this could be improved.

Some form of performance assessment will be found in even the smallest organisation, which makes it pointless to argue whether such assessment is desirable or not. As humans we make judgements, which may or may not be correct, and such judgements rest upon measurement. Much of this performance measurement is informal, as people note the effects of their actions and keep records, possibly for their own benefit.

Nevertheless, there is still considerable concern about the dysfunctional aspects of performance measurement in the public sector. It is well known that the use of performance measures and indicators can have perverse effects and unintended consequences in any organisation. Ridgway (1956), written about 50 years ago

This work is supported by ERSC grant RES-331-25-0005, as part of the UK's Advanced Institute of Management Research.

International Journal of Productivity and Performance Management
Vol. 54 No. 5/6, 2005
pp. 482-493

1741-0401
DOI 10.1108/17410400510604601

discusses this in the very first volume of the *Administrative Science Quarterly*. More recently, Smith (1990) discusses the types of performance indicators (PIs) in use in the public sector and highlights some of the problems that can occur. Smith (1995) discusses some of the known dysfunctional effects of the publication of performance data, taking examples from the UK health sector. Hence, it is clear that performance measurement regimes can have consequences quite different from those intended by their designers. It is also well known (Goldstein and Spiegelhalter, 1996) that league tables can be a very misleading way of presenting performance data, leading to apparent differences in performance when there is no evidence for this. Recognising this, the Royal Statistical Society (Bird *et al.*, 2003) set-up a working party whose report offers advice for improved practice.

Like performance measurement itself, the existence of such perverse effects is uncontroversial and seems to be an accepted part of organisation life. However, these perverse and dysfunctional effects do have costs; some direct and some indirect. These are often ignored when planning a measurement system: presumably in the belief that the system will not produce them. This probably rests on a fondly optimistic conviction that only other people make mistakes. It should surely be obvious that those planning a performance measurement system ought to ensure that its benefits outweigh the costs, and there are two main sets of costs:

(1) The costs of establishing and maintaining the measurement system itself – which will involve staff time, computer resources and other tangible elements. These are relatively straightforward to estimate.

(2) The performance lowering effects of the dysfunctionalities, which may lead to undesirable outcomes and, at the very least, poorer performance than anticipated.

These costs are usually hidden, are rather harder to estimate and are often ignored. Rather than attempting to devise ways to produce accurate estimates of these costs of dysfunctionalities, it makes more sense to try to understand why these perverse effects occur. This should lead to reductions in these costs.

Various forms of balanced scorecard have been introduced in the public sector in recent years, especially in UK government departments. Using scorecards in the public sector is more complicated than in the for-profit sector, since there is rarely a single measure of success. Moore (2003), for example, criticises the use of conventional balanced scorecards in the public and non-for-profit sectors and proposes, instead, a public value scorecard based on the notion of a public value chain. That is, public value rather than some simple notion of cost or profit should drive the scorecard.

The debate about balanced scorecards serves to highlight an important feature of most performance measurement in the delivery of public services: any measurement that focuses solely on a single performance criterion is likely to lead to distortion. Thus, focusing solely on patient waiting times in the health sector may lead to a reduction in other aspects of the quality of a medical intervention. The goals of public service organisations are rarely one-dimensional and this means that any measurement must embody multiple criteria. The representation of performance using approaches such as league tables, star ratings or traffic lights requires the reduction of multiple measures to a one-dimensional summary measure. For example, when evaluating the performance of a hospital, the different performance measures such as waiting times

for surgery and for emergency admission are combined in some way or other. Thus, a general performance indicator P might be calculated as follows;

$$P = w_1 \cdot p_1 + w_2 \cdot p_2 + \cdots + w_n \cdot p_n$$

where P is the overall, summary PI, $p_1, p_2 \ldots p_n$ are the performance scores on the n different measures, and $w_1, w_2 \ldots w_n$ are the weights applied to the individual performance scores.

The relative values given to the ws determines how much of an effect they have on the overall PI. The greater the comparative weight given to each performance score, the greater its effect on the overall indicator. Changing the weights can radically affect the summary indicator. When establishing such a summary indicator, the important questions are: who determines these weights and what values should they take?

The UK performance regime

Performance measurement is often taken to be fundamental to the delivery of improved public services as part of the New Public Management (NPM). There are many definitions of NPM, for example: "a management culture that emphasizes the centrality of the citizen or customer, as well as accountability for results. It also suggests structural or organizational choices that promote decentralized control through a wide variety of alternative service delivery mechanisms, including quasi-markets with public and private service providers competing for resources from policymakers and donors" (Manning, 2000). Since NPM employs a range of service delivery mechanisms, it is hardly surprising that control mechanisms, decentralised or not, and accountability are fundamental to its implementation. Taken to its extreme, NPM can lead to the decentralised and devolved Hollow State in which governments set policy, which is then implemented through managed agencies.

As an example of a public sector performance regime, the UK's current system serves to illustrate the main points of such regimes. Appearing before the UK's Public Administration Select Committee, Michael Barber, who heads the Prime Minister's Delivery Unit (PDMU) described the UK government's performance regime under six headings, as summarised below.

(1) National standards are set-out in Public Service Agreements (PSAs). These specify the goals to which public managers should aspire and towards which they should organise and plan. They are needed so as to ensure that citizens receive a uniformly high standard of public services.

(2) Devolved funding and flexibility is given to units, which, as mentioned above, must organise and plan so as to achieve those targets. Though advice is available on how to achieve the targets, it is left to the service managers to develop and agree delivery plans that specify what will be done.

(3) Regular benchmarking and monitoring of progress using PIs, occurs during the period covered by the PSAs. This allows managers, and the PDMU itself, to see how much progress is being made against planned trajectories and against similar units.

(4) Transfer of best practice is encouraged so as to allow units to learn how best to operate, though local autonomy and specific conditions may mean that the learning is not directly transferable.

(5) The services and the units that deliver them are held to account against the PSA targets, using inspection regimes and published data. The latter are particularly controversial as they may encourage game playing.

(6) Services, units and managers that meet their targets are rewarded for doing so and those that do not may suffer. Thus, hospitals that perform highly against their targets may be given increased autonomy in spending their income.

The rhetoric surrounding this system emphasises that this performance regime and its use of targets must be viewed as a cyclic and iterative processes. Initially, the targets set were fairly crude, but as the departments learn about the process, so it is hoped that the target setting and the use of targets should improve. That is, this is intended as a learning system as well as a control system.

Making things operational

This six-point regime does, it should be noted, allow for different trajectories. That is, they specify the rates of progress likely to be achieved and allow for differences in investment and outcome patterns. For example, the need to retrain staff may mean that it will take at least two years before any improvement is seen, but progress should be rapid thereafter. In such circumstances it would be unreasonable to assess progress against a constant rate of change. The trajectories are specified in the delivery plans that are agreed between the treasury, which provides the funds, and the department(s) responsible for the service in question. The PSA targets are made public and can be viewed on the HM treasury web site, whereas the delivery plans are not.

One obvious problem with a regime of this type is that a government department will have several different targets to meet for its PSA, each of which cascades down through the organisation into a set of lower-level targets. Thus, the number of targets can grow rather large and this brings a couple of other problems in its wake. The first is that of cognitive overload, since Miller's rule of 6 ± 2, seems generally applicable. That is, most people find it almost impossible to concentrate on more than this number of independent variables. There is therefore a tendency to concentrate on particular targets at the expense of others. Linked to this is the public nature of the targets and the bad publicity that this can bring if particular targets are not being met. In response, the PMDU uses a four-level categorisation of problems, as follows.

(1) Something simple, such as a missed deadline or milestone against the planned trajectory.

(2) When there is a problem but the department in question is unsure what to do and asks the PDMU to help. In this case, the latter act as consultants.

(3) This is the first serious level and requires the prime minister's personal involvement in meetings to try to move difficult issues along. This is useful when a problem is shared among two or more government departments.

(4) This is when something is very seriously wrong as shown when published data shows things to be getting worse rather than improving. In such high profile cases, the Prime Minister may make public statements and take the lead.

The categorisation is intended to reflect the intractability of a problem rather than its importance in the PSA universe.

Dysfunctionalities

As mentioned earlier, Smith (1995) lists eight unintended consequences of publishing performance data in the public sector. It is important to realise that this is not an argument against performance measurement per se, more an indictment of the clumsy way that such schemes are sometimes implemented. In a way they reflect Goodhart's Law, which was originally stated in the context of a paper discussing the control of the money supply (Goodhart, 1975) as "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes". It can be stated more generally as something like, "targets are only useful as long as you do not use them to manage by", (Caulkin, 2004). Smith's unintended consequences are as follows.

- Tunnel vision: which occurs when service managers, faced with many different targets, choose the ones that are easiest to measure and ignore the rest.
- Sub-optimisation: when service managers choose to operate in ways that serve their own operation well but damage the performance of the overall system.
- Myopia: when, for whatever reason, managers focus their efforts on short-term targets at the expense of longer-term objectives.
- Measure fixation: when outcomes are difficult to measure there is a natural tendency to use PIs based on measurable outputs. Measure fixation occurs when the PI itself becomes the focus rather than the desired outcome.
- Misrepresentation: this is a form of fraud and occurs when performance data is either misreported or distorted to create a good impression.
- Misinterpretation: this is most evident in performance league tables. The statistical measures are imprecise (Bird *et al.*, 2003; Goldstein and Spiegelhalter, 1996), which means that there may be no real difference between many of the units sequenced in the table – though this may not be obvious from the single-point estimates used.
- Gaming: this occurs when a canny manager deliberately under-achieves in order to secure a lower target in the next round of activity.
- Ossification: which happens when a PI is past its sell-by date and has lost its purpose, but no one can be bothered to revise or remove it.

These dysfunctionalities occur because targets and PIs are used inappropriately, which was the focus of much of the questioning from the Public Accounts Committee (op cit) in 2003. The question is, though, what can be done to avoid these well-known problems?

Virtualism

Though some dysfunctional effects occur because of malpractice, there are other cases in which people strive to be honest and to do a good job and yet these perverse effects still occur. That is, these perverse effects can occur despite the best efforts and good intentions of those developing the measurement systems. Miller (2003) is an anthropologist who studied some UK Local Authorities being inspected by the Audit Commission as part of the Government's Best Value (BV) programme. BV is an improvement regime for which performance measurement and comparison are both fundamental. The main aspects of BV are often summarised as the 4Cs.

- *Challenge*: Each local authority is required to set goals that will ensure that its services are demonstrably in the top 25 per cent of all local authorities.
- *Competition*: All services should be open to competition from other providers, including the private sector.
- *Consultation*: Local authorities are required to seek public opinion of their services.
- *Comparison*: Local authorities must evaluate the quality of their services using quantitative PIs so that they can benchmark themselves against other authorities.

Miller (2003) uses the concept of "virtualism" – when an intended aspect or agent is subtly replaced by another – to describe what he observed. Four types of virtualism are identified in BV and two of these are relevant to this discussion of the unintended consequences of performance measurement. First, some unintended consequences occur because the pursuit of high scores on PIs may displace efforts to improve the actual outcomes. That is, "there is an exaggerated respect for hard quantitative over soft qualitative data" (Miller, 2003, p. 72) – similar to "tunnel vision" as discussed in Smith (1995). Thus, though an individual or organisation may achieve high scores on a PI, this good performance may be virtual rather than real when considered against what the organisation has been established to achieve. In addition, organisations will, if subject to such regimes, employ staff charged to ensure that performance data and reports are presented in ways that place the organisation in the best possible light. Performance presentation becomes a profession.

The second relevant form of virtualism occurs when one agent is subtly replaced by another. It is often said that performance data should be published so that the public may choose among different public service offerings. With this in mind, BV requires local authorities to publish performance data and to engage in public consultation. However, it is very difficult to organise proper public consultation, especially as there is little evidence that the public is hungry for such performance information (Marshall *et al.*, 2000). Persuading members of the public to attend consultation meetings is notoriously difficult and there is a risk that such meetings are dominated by unrepresentative pressure groups. Most local authorities do their best to organise some form of public consultation, but it is often unsatisfactory and unrepresentative. Knowing this, Miller (2003) reports that BV auditors slipped into a role in which they acted as a virtual public. That is, they attempted to assess likely public reactions, because the public itself is uninterested in the performance data.

Why measure performance in public services?

Speaking for the Royal Statistical Society, Bird *et al.* (2003) suggests that the performance of public services should be measured for three reasons:

(1) To establish what works; that is, to use quantitative indicators as a foundation for evidence-based policy by measuring and comparing the performance of different delivery and policy options.

(2) To identify functional competences: which is a step beyond measurement to see what works, since the intention is to identify good performers (and, by implication, poor performers). Often the resulting performance data are

published; for example in school league tables and in the star ratings of NHS hospital trusts and, more recently, the performance of social service departments in UK local authorities.

(3) To support public accountability: which seems to have two dimensions. First elected politicians must demonstrate that the services provided are being delivered properly and represent value for money. Secondly, it is sometimes argued that performance data should be published to enable the public to choose which services to use, when there are alternatives available. This rests on two assumptions: that people wish to choose and that they make use of published performance data to do so. Both of these are questionable.

Measurement for control

Though Bird *et al.* (2003) is surely correct to list these three reasons for control, there is clearly something missing. The use by Michael Barber of words like targets and trajectory is revealing and it is hard to avoid the thought that an important purpose of this measurement is control, as depicted in Figure 1. This shows the usual cybernetic notion of feedback control in which outputs are sampled, compared with a given target value and the difference between the two is then fed back to the input side where corrective action is taken. Thus some value is kept under control and action is taken to ensure that, through time, the variable in question approaches its target value. This cybernetic control works well in physical systems even when there is some turbulence. However, as Hofstede (1981) points out, this notion of control rests on some important assumptions.

- The objectives are wholly unambiguous and are agreed by everyone involved. That is, there is no uncertainty about values and complete agreement amongst stakeholders.
- Outputs from the system can be usefully quantified and measured.
- The effects of taking action and intervening are known: that is, there is sufficient knowledge of the way the system operates that it is clear what action should be taken when outputs deviate from the target. Clearly this, in turn, depends on unambiguous targets and some quantification.
- The activity being controlled is repetitive so that learning may occur. This means that, though the effects of intervention may not be fully understood, this understanding may grow through time.

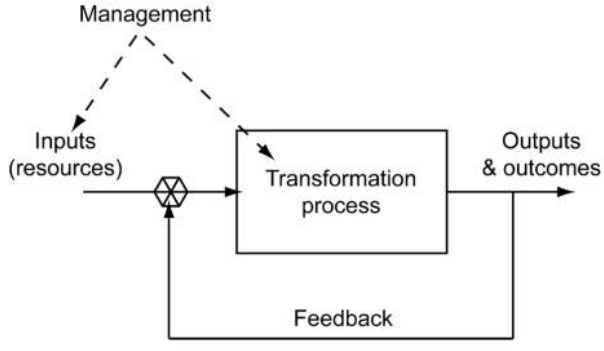

Figure 1.
The cybernetic control metaphor

If all these assumptions hold, then Hofstede (1981) argues that routine control is possible.

How do these assumptions fit the delivery of public services? Figure 2 shows a spectrum of types of service. On the left hand side are departments such as the Inland Revenue and Work and Pensions in which most of the staff are low paid and relatively unskilled people, whose work involves following tightly defined protocols. The staff have limited discretion and are, in some sense, interchangeable. By contrast, the right hand side is typified by health and education in which people delivering the service are highly qualified professionals who are, relatively, well paid. As professionals, they have considerable scope to vary the protocols that define their work and are expected to reflect on their practice so as to improve it.

It is tempting to argue that work on the right hand end of the spectrum should be reduced in scope and variety, since then it can be treated as if it were like that on the left hand end. Some measurement regimes are indeed designed to have this effect and, as a response to this, Noordegraaf and Abma (2003) argues that attempts to reduce uncertainty and ambiguity may be mistaken. This conclusion is reached from a categorisation of public management practices into three groups:

(1) *Canonical practices.* Those in which all the important issues are known, the necessary standards are uncontested and it is clear how to address the issues. It is sensible to use quantitative PIs to manage the performance of such practices and ambiguity and uncertainty should, as far as possible, be reduced.

(2) *Practices in transition.* These apply in two situations. First, when facing complex issues that are difficult to address even though they are known and understood and even when values and standards are uncontested. For example, the necessary action may be just too disruptive. Secondly, they occur when issues are known and understood, but standards are contested – which leads to a debate about values and a need to mobilise support. In these practices in transition, quantitative PIs are of limited value and assessment is best done by expert opinion based on negotiated professional agreement.

(3) *Non-canonical practices.* These apply in processes that are non-routine, fuzzy, innovative and conflictual, that is in situations considered to be wicked problems (Rittel and Webber, 1973). In such situations, the conventional assumptions about measurement systems and control make little sense. This does not mean that measurement has no value, but does mean that great care is needed to avoid virtualism.

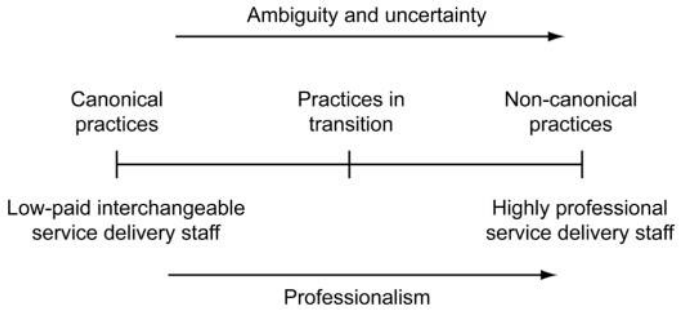

Figure 2.
Ambiguity, uncertainty and non-canonical practices

The mistake, so often, is to apply performance management approaches that are appropriate to canonical practices to those that are not: that is, to practices in transition, and to non-canonical practices. It is clear that all three types of practices are found in the delivery of public services and that ambiguity is sometimes necessary if any action is to be taken. Closing down this ambiguity may make action impossible.

In the terms of Hofstede (1981), performance measurement can easily be used for routine control at the left hand side of Figure 2 whereas it may only be applicable for the simplest aspects of the work of staff at the right hand end. This suggests that applying the methods appropriate to routine control to work at the right hand end could be a serious mistake.

Another way of thinking about this

Hood (1999) uses ideas from cultural anthropology to discuss different approaches to public administration and management so as to tease out what is significantly about NPM. Hood's argument is based on a simple 2 × 2 typology (Figure 3), in which the dimensions and categories are a simplified version of those proposed by Douglas (1982). The typology was originally intended by Douglas to explain how individuals in different societies related to one another and it assumes that culture (people's shared beliefs and norms) is based on particular patterns of social relations. These in turn are measured on two dimensions:

(1) *Grid*. This indicates the degree to which people's actions are governed by externally imposed rules and conventions. If grid is high, there is little scope for individuals to negotiate what they do and how they do it.

(2) *Group*. This indicates the degree to which actions are governed by group choice that is the social collectivity to which someone is committed.

Amongst other things, Hood (1999) uses this typology to discuss four different approaches to control, though devotes little space to performance measurement. Here it is used to think about dysfunctionality that may result from performance measurement.

The four categories of the typology are:

(1) *Hierarchist*. This is a highly prescribed world in which external rules and group norms tightly define what is permissible. Thus, personal security and predictability are highly valued.

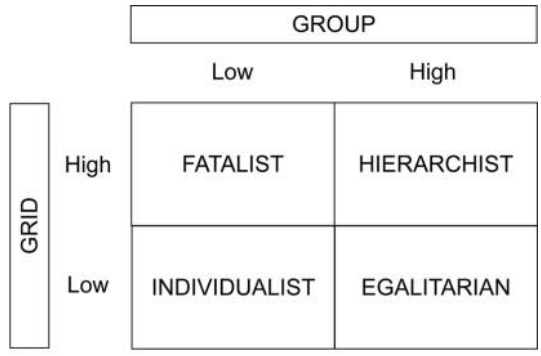

Figure 3.
Simplified grid: group theory

(2) *Individualist.* This lies on the same diagonal as hierarchist and is therefore a world in which individuals have great freedom of choice but may have little security.

(3) *Fatalist.* In this world, the individual is placed in tightly prescribed roles with little or no autonomy and no obvious way to cooperate with others in the group.

(4) *Egalitarian.* In this world there are no clear rules defining the roles of members, but there is a clear distinction between those who are part of the group and those who are not. Active participation in group activities is expected.

This grid:group typology is helpful in understanding the limits of externally imposed performance measurement and indicates why some of the perversities and dysfunction occurs. It should be obvious that such measurement and its implicit or explicit control are best fitted to worlds in which the grid dimension is high. If, in turn, the group dimension is high then this indicates a highly conformist world in which deliberate perversion of performance measurement is unlikely – at least from an internal perspective.

Grid:group and performance measurement

Figure 2 showed a spectrum of service types provided in the public sector and it is interesting to try to relate this to grid:group cultural theory. The left hand end of Figure 2 depicts services delivered by low-paid, interchangeable staff whose work consists of the application of tightly defined rules and protocols. It seems that the culture of an organisation whose members are of this type is likely to be closest to that of the fatalist or hierarchical positions in Figure 3. An important point of grid:group theory is the idea that groups cannot occupy multiple positions in the typology, but tend to move towards the centre of one of the four locations. If the service is delivered by staff with little or no opportunity to interact with one another, then the fatalistic position applies. On the other hand, if group norms strongly reinforce the externally defined protocols, then a hierarchical position will apply.

The right hand end of Figure 2 shows services of the type delivered by highly professional and, probably, well-paid staff. Given their professional status it should be clear that these organisations would score highly on the group dimension – to some degree, all professions are self-regulating and define their own norms for behaviour. What of the grid dimension? If the staff also work to externally defined norms and procedures, then the hierarchist position seems the best way to regard them. On the other hand, if staff are left to define their own work and standards, then an egalitarian position is occupied. It seems most likely that a hierarchist position will apply in the public sector.

Thus, most services in the UK public sector are likely to be delivered by organisations with cultures that are, in the terms of grid:group theory, primarily fatalist or hierarchist. Does this provide any insight into what may go wrong with performance measurement? To answer this question, we must return to the list of unintended consequences listed in Smith (1995).

Understanding unintended consequences

When an organisation or part of it becomes fatalistic but is treated as if it were hierarchical, what forms of dysfunctionality are likely to result from the publication of

performance data? Particularly, what will happen if the publication of these data is enforced from on high with little or no local autonomy?

First and perhaps most obvious is tunnel vision, in which people focus on things that are easiest to measure. Why should they bother to do anything else, if there is little reinforcement from their group but just abstract rules provided from on high? The same argument might be made about sub-optimisation, operating in ways that serve their own operation well but damage the performance of the overall system. The same is probably true of myopia (focusing on the short-term) and measure fixation, which the focus is measurable outputs rather than desired outcomes. The most extreme case is that of misrepresentation to create a good impression and a subtler version is found in gaming to ensure an easier life in the future.

Smith's other two categories of misinterpretation and ossification perhaps apply higher up the hierarchy and cannot be analysed in these terms. However, it seems that one way to understand some of the most common unintended consequences of publishing performance data is to realise that fatalistic behaviour may occur in organisations that those at the top assume to be hierarchist. There is obviously scope for further research.

References

Bevan, G.M. and Hood, C. (2004), "Targets, inspections, and transparency", *BMJ*, Vol. 328, p. 598.

Bird, S.M., Cox, D., Farwell, V.T., Goldstein, H., Holt, T. and Smith, P. (2003), *RSS Working Party on Performance Monitoring in the Public Services*, Royal Statistical Society, London.

Caulkin, S. (2004), "Against the grain", *Society Guardian*, 31 March.

Douglas, M. (1982), "Cultural bias", *The Active Voice*, Routledge, London.

Goldstein, H. and Spiegelhalter, D.J. (1996), "League tables and their limitations: statistical issues in comparisons of institutional performance", *Journal of the Royal Statistical Society, A*, Vol. 159, pp. 385-443.

Goodhart, C.A.E. (1975), "Monetary relationships: a view from Threadneedle Street", *Papers in Monetary Economics*, Reserve Bank of Australia, Vol. I.

Hofstede, G. (1981), "Management control of public and not for profit activities", *Acctg, Orgs and Soc*, Vol. 6 No. 3, pp. 193-211.

Hood, C. (1999), *The Art of the State: Culture, Rhetoric and Public Management*, Oxford University Press, Oxford.

Manning, N. (2000), "The New Public Management & its legacy", available at: www1.worldbank.org/publicsector/civilservice/debate1.htm

Marshall, M.N., Shekelle, P.G., Leatherman, S. and Brook, R.H. (2000), "The public release of performance data. What do we expect to gain? A review of the evidence", *The Journal of the American Medical Association*, Vol. 283, pp. 1866-74.

Miller, D. (2003), "The virtual moment", *Royal Anthropological Institute*, Vol. 9 No. 1, pp. 57-75.

Moore, M.H. (2003), "The public value scorecard: a rejoinder and an alternative", review of strategic performance measurement and management in non-profit organizations, by Robert Kaplan, HCNO working paper series, May.

Noordegraaf, M. and Abma, T. (2003), "Management by measurement? Public management practices amidst ambiguity", *Pub. Admin.*, Vol. 81 No. 4, pp. 853-73.

Ridgway, V.F. (1956), "Dysfunctional consequences of performance measurements", *Admin. Sci. Qtly*. Vol. 1 No. 2, pp. 240-7.

Rittel, H.W.J. and Webber, M.M. (1973), "Dilemmas in a general theory of planning", *Policy Sci.*, Vol. 4, pp. 155-69.

Smith, P.C. (1990), "The use of performance indicators in the public sector", *Journal of the Royal Statistical Society, A*, Vol. 153 No. 1, pp. 53-72.

Smith, P. (1995), "On the unintended consequences of publishing performance data in the public sector", *Intl. Jnl. Pub. Admin.*, Vol. 18 Nos. 2/3, pp. 277-310.

Stationery Office (2003), "On target? Government by measurement", Public Administration Select Committee, Fifth report, London (HC 62-I.).

This article has been cited by:

1. Max Visser Management Control, Accountability, and Learning in Public Sector Organizations: A Critical Analysis 75-93. [Abstract] [Full Text] [PDF]
2. Sarah Holzapfel. 2016. Boosting or Hindering Aid Effectiveness? An Assessment of Systems for Measuring Donor Agency Results. *Public Administration and Development* **36**:1, 3-19. [CrossRef]
3. Rob Schoenmaker Faculty Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands Hans de Bruijn Faculty Technology, Policy and Management, Delft University of Technology, Delft, Netherlands . 2016. Embracing complexity in performance-based contracts for road maintenance. *International Journal of Productivity and Performance Management* **65**:1, 4-24. [Abstract] [Full Text] [PDF]
4. Daniela Monacelli, Aline Pennisi. 2015. Spending Review: una, nessuna, centomila. *ECONOMIA PUBBLICA* :1, 69-108. [CrossRef]
5. CHRISTINA BOSWELL. 2015. THE DOUBLE LIFE OF TARGETS IN PUBLIC POLICY: DISCIPLINING AND SIGNALLING IN UK ASYLUM POLICY. *Public Administration* **93**:2, 490-505. [CrossRef]
6. Joseph Drew, Brian Dollery. 2015. Inconsistent Depreciation Practice and Public Policymaking: Local Government Reform in New South Wales. *Australian Accounting Review* **25**:1, 28-37. [CrossRef]
7. Vicki Johansson. 2015. When will we ever learn ?. *NISPAcee Journal of Public Administration and Policy* **8**:2. . [CrossRef]
8. Steve Martin James Downe Performance Measurement of Local Public Service Networks 43-64. [Abstract] [Full Text] [PDF]
9. David Mason, Carola Hillenbrand, Kevin Money. 2014. Are Informed Citizens More Trusting? Transparency of Performance Data and Trust Towards a British Police Force. *Journal of Business Ethics* **122**:2, 321-341. [CrossRef]
10. Frank Ohemeng, Elyse McCall-Thomas. 2013. Performance management and "undesirable" organizational behaviour: Standardized testing in Ontario schools. *Canadian Public Administration* **56**:3, 456-477. [CrossRef]
11. John HoodGlasgow, UK Tom SmithMarple Bridge, UK. 2013. Perceptions of quantifiable benefits of local authority risk management. *International Journal of Public Sector Management* **26**:4, 309-319. [Abstract] [Full Text] [PDF]
12. John R. MansfieldCloister Green Research, Canterbury, UK. 2013. Heritage protection in England: the New Labour legacy. *Structural Survey* **31**:1, 6-20. [Abstract] [Full Text] [PDF]
13. Maria Manuela PereiraEuropean School of Luxembourg I, Luxembourg Nuno Filipe MelãoCentro Regional das Beiras, Universidade Católica Portuguesa, Viseu, Portugal. 2012. The implementation of the balanced scorecard in a school district. *International Journal of Productivity and Performance Management* **61**:8, 919-939. [Abstract] [Full Text] [PDF]
14. S. Guilfoyle. 2012. On Target?--Public Sector Performance Management: Recurrent Themes, Consequences and Questions. *Policing* **6**:3, 250-260. [CrossRef]
15. Deryl NorthcottDepartment of Accounting, AUT Business School, The Auckland University of Technology, Auckland, New Zealand Tuivaiti Ma'amora TaulapapaManukau Institute of Technology, Auckland, New Zealand. 2012. Using the balanced scorecard to manage performance in public sector

organizations. *International Journal of Public Sector Management* **25**:3, 166-191. [Abstract] [Full Text] [PDF]

16. Gerald Eisenkopf. 2011. Paying for better test scores. *Education Economics* **19**:4, 329-339. [CrossRef]
17. David A. BuchananCranfield University School of Management, Cranfield, UK John StoreyThe Open University Business School, Milton Keynes, UK. 2010. Don't stop the clock: manipulating hospital waiting lists. *Journal of Health Organization and Management* **24**:4, 343-360. [Abstract] [Full Text] [PDF]
18. Jennifer ColeUniversity Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK Zoe RadnorWarwick Business School, University of Warwick, Coventry, UK. 2010. How healthy is the annual health check?. *International Journal of Health Care Quality Assurance* **23**:6, 537-553. [Abstract] [Full Text] [PDF]
19. Mahmoud M. YasinDepartment of Management and Marketing, East Tennessee State University, Johnson City, Tennessee, USA Carlos F. GomesSchool of Economics, University of Coimbra, ISR-Institute of Systems and Robotics, Coimbra, Portugal. 2010. Performance management in service operational settings: a selective literature examination. *Benchmarking: An International Journal* **17**:2, 214-231. [Abstract] [Full Text] [PDF]
20. Eliana Minelli, Gianfranco Rebora, Matteo Turri. 2009. Why do controls fail? Results of an Italian survey. *Critical Perspectives on Accounting* **20**:8, 933-943. [CrossRef]
21. Karen FryerCaledonian Business School, Glasgow Caledonian University, Glasgow, UK Jiju AntonyDepartment of DMEM, Strathclyde Institute for Operations Management, University of Strathclyde, Glasgow, UK Susan OgdenCaledonian Business School, Glasgow Caledonian University, Glasgow, UK. 2009. Performance management in the public sector. *International Journal of Public Sector Management* **22**:6, 478-498. [Abstract] [Full Text] [PDF]
22. Gerald Eisenkopf. 2009. Negative Weights for Performance Measures. *International Public Management Journal* **12**:3, 332-344. [CrossRef]
23. Mervyn ConroyHealth Services Management Centre, University of Birmingham, Edgbaston, UK. 2009. Narrative-ethics informed approach to health services leadership development. *Leadership in Health Services* **22**:3, 259-272. [Abstract] [Full Text] [PDF]
24. Simmy M. MarwaBased at the Hamdan Bin Mohamed E-University, Dubai, United Arab Emirates Mohamed ZairiBased at the European Centre for Total Quality Management (ECTQM), School of Management, University of Bradford, Bradford, UK. 2009. In pursuit of performance-oriented civil service reforms (CSRs): a Kenyan perspective. *Measuring Business Excellence* **13**:2, 34-43. [Abstract] [Full Text] [PDF]
25. Daryl MayRachel MacdonaldSheffield Hallam University, Sheffield, UK Ilfryn PriceSheffield Hallam University, Sheffield, UK Phil AskhamSheffield Hallam University, Sheffield, UK. 2009. Excellent patient environments within acute NHS trusts. *Journal of Facilities Management* **7**:1, 7-23. [Abstract] [Full Text] [PDF]
26. Ilfryn PriceFacilities Management Graduate Centre, Sheffield Hallam University, Sheffield, UK Elizabeth ClarkFacilities Management Graduate Centre, Sheffield Hallam University, Sheffield, UK. 2009. An output approach to property portfolio performance measurement. *Property Management* **27**:1, 6-15. [Abstract] [Full Text] [PDF]
27. Eliana Minelli, Gianfranco Rebora, Matteo Turri. 2008. How Can Evaluation Fail? The Case of Italian Universities. *Quality in Higher Education* **14**:2, 157-173. [CrossRef]

28. Zoe Radnor. 2008. Muddled, massaging, manœuvring or manipulated?. *International Journal of Productivity and Performance Management* **57**:4, 316-328. [CrossRef]
29. Eliana MinelliUniversity Carlo Cattaneo – LIUC, Italy Gianfranco ReboraUniversity Carlo Cattaneo – LIUC, Italy Matteo TurriDepartment of Economics, Business and Statistics, University of Milan, Italy. 2008. The risk of failure of controls and levers of change: an examination of two Italian public sectors. *Journal of Accounting & Organizational Change* **4**:1, 5-26. [Abstract] [Full Text] [PDF]
30. Mike Bourne, Steven Melnyk and Norman FaullRichard GreatbanksDepartment of Management, School of Business, University of Otago, Dunedin, New Zealand David TappDunedin City Council, Dunedin, New Zealand. 2007. The impact of balanced scorecards in a public sector environment. *International Journal of Operations & Production Management* **27**:8, 846-873. [Abstract] [Full Text] [PDF]
31. Dries Verlet, Carl DevosBefore the Mining Begins 1-20. [CrossRef]