Lancaster University Management School
Working Paper
2005/024

A call for conceptual clarity: A soft systems view of performance measurement in public service delivery

Mike Pidd

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

A CALL FOR CONCEPTUAL CLARITY: A SOFT SYSTEMS VIEW OF PERFORMANCE MEASUREMENT IN PUBLIC SERVICE DELIVERY

Michael Pidd
Lancaster University Management School
& Advanced Institute of Management Research
Lancaster University
Lancaster LA1 4YX

M.Pidd@lancaster.ac.uk

ABSTRACT

Performance measurement systems and the publication of performance data are fundamental to the New Public Management, with its emphasis on decentralised service provision through a variety of agencies. There seem to be four reasons for this performance measurement: to see what works, to identify competences, to support public accountability and to allow control of decentralised service provision. Each of these is examined, using the root definitions that form part of soft systems methodology, which highlights the importance of different worldviews that provide justifications for different approaches to this measurement. If performance measurement systems are to be beneficial, their design should be based on conceptualisations that recognise these different worldviews and purposes.

KEYWORDS

Performance measurement, public sector, soft systems methodology

INTRODUCTION

The last twenty years have seen major changes in the ways that some governments provide services to their publics. In some countries, notably the UK and New Zealand, there has been a shift from service provision via classic bureaucratic organisations to delivery via decentralised agencies and private sector providers. This latter mode of operation is often taken as a major part of the New Public Management (NPM). There are many definitions of NPM, but the following two are sufficiently rich to capture its essence:

- Manning (2003), on the World Bank web-site, offers the following: "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."
- Gow and Dufour (2000), who discuss whether NPM can be properly considered as a paradigm, suggest that: "There are two main ideas when introducing NPM in public administration: (1) the separation of policy formulation from operations; and (2) the importance of management, inspired by private sector management." (citing Charih and Rouillard, 1997: 27)

This paper is not concerned with the degree to which NPM does or does not mimic the management of private sector organisations. Rather, its concern is the different uses made of performance measurement in public services when their actual provision is separated from policy and when a range of delivery mechanisms is in use – both being features of NPM.

There are many different ways in which performance measurement can form part of NPM. Perhaps its most developed form is found in the UK government's use of Public Service Agreements (PSAs), which are service level agreements between the Treasury and the service providers. The workings of the PSA system were the subject of discussions in the UK's Parliamentary Select Committee on Public Administration (Stationery Office, 2003). From this, it seems that the main components of the PSA system are as follows.

1. As a result of discussion about expenditure between service providers and the Treasury, national standards are set out in Public Service Agreements (PSAs). Once the spending round is complete, these PSAs are made public and are intended as goals to which

public managers should aspire and for which funding will be provided. For example, the 2004 PSA of the Home Office includes the following (HM Treasury, 2004).

- Reduce crime by 15%, and further in high crime areas, by 2007-08.
- Reassure the public, reducing the fear of crime and anti-social behaviour, and building confidence in the Criminal Justice System without compromising fairness.

Note that one of these is stated in quantitative terms and the other is not.

2. Once the PSAs are established, together with the funding to achieve them, the delivery units are allowed flexibility in how they organise and plan to achieve them. The resulting Delivery Plans, though published, are intended for use by service managers to specify what actions will be taken to achieve the targets. That is, Delivery Plans are intended for internal use as part of the process of planning to achieve targets and in assessing progress against those targets.
3. Performance indicators, usually quantitative, are fundamental to this progress monitoring, as is the use of comparative benchmarking. Their use allows progress to be monitored against planned trajectories and against similar units.
4. As an extension of benchmarking, there are attempts to transfer best practice, if this is appropriate, by seminars and other meetings at which managers present their work.
5. There is external auditing of progress against the PSA targets, using inspection teams and published data. Performance data is regularly published and taken up by the media, who are wont, correctly or not, to construct league tables.
6. Services, units and managers that meet their targets are rewarded for doing so and those that do not may suffer. Thus, for example, hospitals that perform well against their targets may be given increased autonomy in managing their finances and planning their services. Units that do not may be labelled as failing and managers may lose their jobs.

Listed so tersely, such a performance regime can seem rather simple-minded. However, this would be a very unfair conclusion for it is important to realise that the six points are set within an iterative framework in which it is recognised that target setting and measurement are not straightforward. Thus, targets and measures may be initially rather crude, but there is an expectation that the performance of this performance management system will itself improve after several iterations. There are central units, some housed in the Treasury and a Prime Minister's Delivery Unit, all tasked with ensuring that this performance regime is implemented effectively.

This brief description reveals a tightly organised performance management system that is rather different from traditional bureaucratic public service organisation. Other countries, notably New Zealand, have followed similar routes with similar aspirations, though obviously the detail is somewhat different. However, the above list may be taken as typical of the aspirations for performance measurement and management as employed in the delivery of public services. National targets are set, which then cascade down to local targets for service provision. Carrots and sticks are used to encourage the achievement of these targets, including increased autonomy for success, and naming and shaming for failure. It is well-known that, though powerful, such approaches have severe limitations and that alternative approaches may be preferable in some circumstances (Smith, 1995; Pidd, 2005).

MEASURING PERFORMANCE

Performance measurement is pervasive in organisations, whether public or private, and it is pointless to argue whether it is desirable or not. Some measurements are informal, as people keep their own records to see how particular activities are progressing. Other measurements are much more formal and may be required by law – for example, financial accounts must be produced, audited and published so as to reduce the possibility for fraud and to allow stakeholders to assess how well things are going. Such financial measurement is taken for granted in for-profit organisations and is also pervasive, in slightly different forms, in the public and not-for-profit sectors.

The last 20 years have seen increasing interest in the measurement of other aspects of performance. The most widely used approach seems to be balanced scorecards, as proposed by Kaplan and Norton (1996). In their original form, these balanced scorecards were attempts to encourage businesses to focus on more than the immediate financial bottom line. They suggested that organisations attempt to measure their performance on four different dimensions:

1. *Learning and growth*: how well does the business develop its people for the futures? A long term view
2. *Business processes*: how well does the business operate? An internal view.
3. *Customers*: how does the business appear to its customers? An external view.
4. *Financial*: how well does the business perform in financial terms? A current view.

Such a list seems sensible, but it raises the question of how managers should decide what to measure within these four dimensions? Financial measurement is well-established, at least in terms of current performance, but what of the other dimensions? The usual advice is to focus on those aspects of performance in the other three dimensions that are leading indicators of future financial performance. This rests on a sensible assumption that, unless a business develops its assets, organises itself efficiently and looks after its customers, its future financial performance is likely to be poor. However, it also suggests that, even when balanced scorecards are used, the financial bottom line has the strongest pull in the private sector.

Balanced scorecards seem to be ubiquitous in the UK public sector too, with considerable sums spent on external consultants to ensure their implementation. It is clear, however, that they should take a different form to those used in the for-profit sector in which bottom-line financial concerns predominate. In public service provision, the aim is not to maximise profit or returns on investment; that is, finance is a means to an end and not an objective. Hence, critics have suggested that public sector scorecards should be rather different. Moore (2003), for example, suggests a public value scorecard based on a public value chain in which public value replaces the financial bottom-line as the driver for the various measures included in the scorecard. Scorecards can be regarded as a welcome recognition that any measurement of performance, certainly in the provision of public services, is multi-dimensional. Rather than hiding this multi-dimensionality away in a summary measure, the various criteria and measures are displayed alongside one another, leaving for debate the question of which criteria are most important.

WHY MEASURE THE PERFORMANCE OF PUBLIC SERVICES?

The Royal Statistical Society (RSS) is much exercised about performance measurement in public services, this having been a major concern since its foundation. For example, in one of the RSS journals, Goldstein and Spiegelhalter (1996) strongly criticises the way in which published performance indicators are often unthinkingly used in constructing league tables. The basis of the critique is that performance indicators are, in essence, based on simple statistical models and that any resulting performance estimates should carry reasonable estimates of error. Once these are included in the measures, then much of the apparent ranking in league tables disappears, since there is no statistically significant difference in performance between the many of the units ranked in the tables. As in most sports, mid-table teams are all pretty much of a muchness. This and other examples caused the RSS to call for more appropriate performance measurement in public services, preferably on some scientific basis.

Hence it commissioned a report (Bird et al, 2003) in which the various problems are discussed and proposals are made for improvement.

Reasonably enough, Bird et al (op cit) addresses the question: why measure performance? It proposes three reasons:

1. *To see what works*: this is the use of quantitative indicators as a foundation for evidence-based policy by measuring and comparing the performance of different delivery and policy options. To be properly done, this requires the careful use of statistical methods and designed comparisons: some of the possible approaches are described in Boyne et al (2002). Thus, if there are several options available, the performance of each can be measured and this information can be used for comparison to determine the best way to proceed. Alternatively, the impact of a single policy might be measured by comparing its costs and benefits to see if it is worthwhile.
2. *To identify competences*: this is the use of quantitative indicators to identify good performers (and, by implication, poor performers). Often the resulting performance data are published; for example in school league tables, the star ratings of NHS hospital trusts and, more recently, the performance of social service departments in UK local authorities. An ultimate aim of such measurement is to encourage the transfer of knowledge and expertise and to inform services managers about how well they are performing. The idea is that the identification of high performers will enable poorer performers to learn how to improve – it also provides positive feedback to the high performers.
3. *To support public accountability*: this rests mainly on the publication of performance data to allow members of the public to see whether services are being delivered properly and offer value for money. Since NPM separates policy from operations, it is clear that the accountability loop must be closed in some way; otherwise there will be no link between policy and action. Though this measurement may, in theory, be for public use, politicians are intimately concerned with this aspect of measurement since part of their future may depend on it. Whether the general public is so interested is a moot point. The indications from local authorities (Miller, 2003) and health care (Marshall et al, 2000) suggest that public interest is rather limited.

It should, though, be clear that this three-part list is incomplete: that is, something is missing. From the earlier description of the UK's PSA system it is clear that retaining central control with decentralised delivery is a major concern and forms a major part of its rationale. Hence, there is clearly a fourth reason for measuring the performance of public services: to enable central control. That is, performance indicators are used to see what works, to identify competences, to support public accountability and to allow control of decentralised service provision. Having four different aims, allows plenty of space for confusion and the rest of this paper uses ideas from Soft Systems Methodology to try to bring some conceptual clarity into the continuing discussion about performance measurement in the provision of public services.

SOFT SYSTEMS METHODOLOGY AND ROOT DEFINITIONS

Soft Systems Methodology (SSM) was developed by Checkland and colleagues at Lancaster through an action research programme that ran for over 20 years. For brief accounts of SSM see Checkland (2001) or Pidd (2003, chapter 5) and for the complete account, see Checkland (1981). For an account of a public sector study that was based on SSM, see Brown et al (2005) or Brown and Cooper (2004).

Defining what we mean by system

The word 'system' is in everyday use and some have argued for many years that we live in a 'systems age' (Ackoff, 1974). People commonly refer to human nervous systems, to taxation systems, to transport systems and, in the context of this paper, to performance measurement systems. However, this everyday use of the term 'system' can be very misleading and it is important develop a careful definition so as to avoid unnecessary confusion. This is particularly important when applying SSM, in which a system is a concept used to account for behaviour observed or expected in the external world. That is, systemicity lies in the eyes of the beholder.

With this in mind, Checkland (1981, p 317) defines a system as: 'a model of a whole entity; when applied to human activity, the model is characterised fundamentally in terms of hierarchical structure, emergent properties, communication and control. An observer may choose to relate this model to real-world activity. When applied to natural or man-made entities, the crucial characteristic is the emergent properties of the whole'. SSM is particularly concerned with human activity systems – purposive (i.e. serving a purpose) systems that express some purposeful (i.e. deliberate or willed) human activity. That is, a human activity system is a conceptual model of some purpose-designed entity in which humans engage in deliberate activity.

SSM as a methodology

In SSM, a human activity system is an intellectual construct that can provide a useful way of discussing organised human action. There is, though, no need to assume that such purposive systems actually exist, just to accept that they are convenient. As a methodology, SSM is a learning cycle that aims to support people engaged in the messy task of real-world thinking and problem solving. This paper is not the place to provide an extended account of SSM as a methodology for which readers should consult the references cited earlier. Checkland's own description of SSM has varied over the years and the current favourite is shown in figure 1, based on that in Checkland and Holwell (2004, p 52).

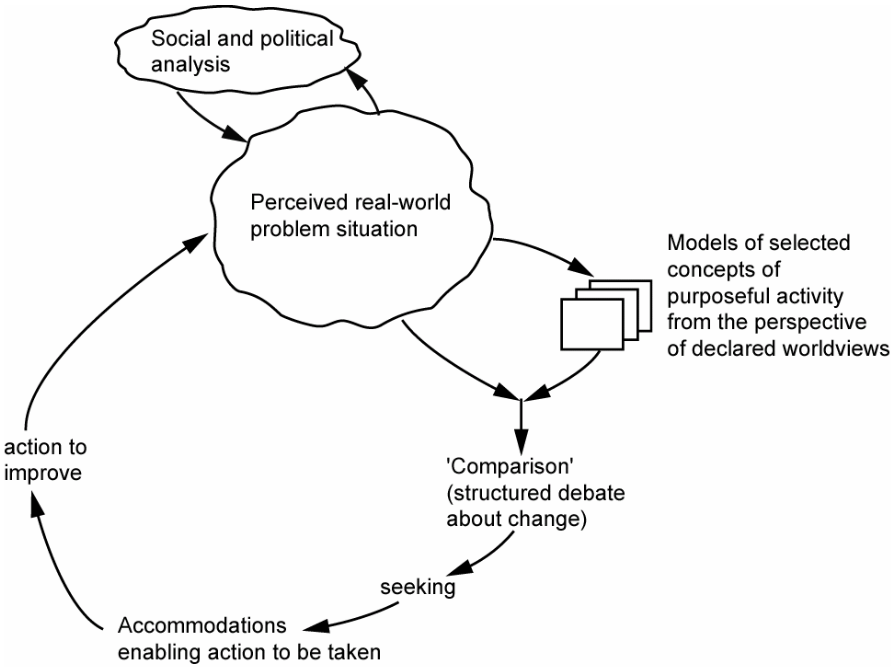

Figure 1: An outline of soft systems methodology

The SSM learning cycle begins with a real-world problem situation, which some people have come to regard as problematic enough to warrant serious attention, usually with a view to seeking improvement. This problem situation needs to be explored, so as to understand the links between its various facets that have led to whatever is regarded as unsatisfactory. Checkland and Scholes (1990) suggests the use of a range of analyses for this investigation so as to lead to the definition of some relevant purposeful activity models. These models are often developed using 'root definitions' – a concept that will be used later in this paper to explore different views of performance measurement. These models of how things could be are then compared with the current situation to support a debate about desirable change. To make progress it may be necessary to repeat the SSM cycle a few times.

Root definitions

The aim of an SSM study is usually to bring about some improvement, however this is defined, and making an improvement usually requires some change to be made. Rather than proceed direct to a redesign of the unsatisfactory human activity system, it is normal to present idealised models, or hypotheses, of how these might be composed. These idealisations are captured in root definitions. Mingers and Taylor (1992) reports on a survey of SSM use that shows root definitions to be among the most commonly employed techniques of SSM.

Root definitions are normally summarised in the CATWOE mnemonic: Customers, Actors, Weltanschauung, Ownership and Environmental Constraints. They are discussed in detail as part of conceptual modelling in Wilson (2001). The following brief account of CATWOEs is taken from Pidd (2003, p125ff).

- *Customers*. These are the immediate beneficiaries or victims of what the system does. It can be an individual, several people, a group or groups. This is very close to the total quality management (TQM) notion that the customer is the next person to receive the work in progress. It indicates what happens to the output from the system and forms part of the external relations of the system.
- *Actors*. In any human activity system there are people who carry out one or more of the activities in the system, these are the actors. They form part of the internal relations of the system. There may be several actors or several groups and their relationships also form part of the internal relations of the system.
- *Transformation process*. This is the core of the human activity system in which some definite input is converted into some output and then passed on to the customers. The actors take part in this transformation process. The process is an activity and its description therefore requires the use of verbs. Ideally, a root definition should focus on a single transformation.
- *Weltanschauung*. This is the, often taken for granted, outlook or world view which makes sense of the root definition being developed. It is important to specify this

because any system definitions can only make sense with some defined context. Thus a root definition needs only a single Weltanschauung.

- *Ownership*. This is the individual or group responsible for the proposed system in the sense that they have the power to modify it or even to close it down. This can overlap with the actors of the system.
- *Environmental constraints*. All human activity systems operate within some constraints imposed by their external environment. These might be, for example, legal, physical or ethical. They form part of the external relations of the system and need to be distinguished from its ownership.

The six elements of a root definition need to be consistent with one another; otherwise the root definition will not make sense. It is particularly important to consider the Weltanschauung, since a root definition will only make sense within this worldview or ideology. Many human activity systems can be conceptualised from a range of worldviews, which is one reason why different people may quite reasonably propose radically different options for change when considering a problem situation. CATWOEs provide a neutral language that permits different worldviews to be teased out during the process of formulating options for change. This can be helpful in highly charged situations in which opinions differ markedly.

Though most 'classical' SSM studies are aimed at improvement, here the idea is to support some thinking about why performance measurement is a part of public service provision under NPM. That is, this paper does not describe an improvement study, rather it uses root definitions as a means to analyse the four different reasons cited earlier for measuring the performance of public services.

CONSTRUCTING ROOT DEFINITIONS FOR PERFORMANCE MEASUREMENT

An earlier section suggested that there are four main reasons for employing performance measurement in the delivery of public services: to see what works, to identify competences, to support public accountability and to allow control of decentralised service provision. Root definitions will be constructed for each of these, to see what light such an analysis can shed.

To see what works

In any root definition it is important to establish the Weltanschauung, or worldview, within which the description make sense. Of course a description could make sense within several worldviews, but their consideration is still a good starting point. In this case, it is perhaps best to assume that performance measurement is part of a drive for evidence-based policy. That is, for ensuring the policy for public service delivery is based on the type of evidence that comes from a performance evaluation. From this, it possible to construct a possible root definition under the six CATWOE headings.

Customers: since the performance evaluation is conducted to inform the policy makers then it seems reasonable to regard them as the main customers. Though it is of course true that the public should benefit from improved policy, the customers in a root definition are the immediate beneficiaries or victims: in this case, the policy makers.

Actors: were the proposals of the RSS, summarised in Bird et al (op cit) to be implemented, then an independent cadre of performance analysts would occupy this role. However, real life is somewhat different and so the best than can be said is that performance measurement to see what works will be run by a group appointed to do so.

Transformation: the aim of such a performance measurement system is to produce proper evidence of whether or not a policy works or is better or worse than some alternative. This must, of course, be done in the light of the purpose of the public service under scrutiny.

Weltanschauung: given the opening sentences of this section, a worldview within which the rest of the root definition makes sense is one which states that policy based on soundly collected and analysed evidence is likely to be better than that based solely on dogma or expediency.

Ownership: there seem to be two communities that can modify or close down such performance measurement systems and these are the political and policy communities as reflected in the government of the day.

Environmental constraints: the obvious ones to consider relate to the legal system, available finances and available skills, which will all directly affect such performance measurement in practice.

Hence, summarising this root definition in a couple of sentences leads to something like the following: *A system that provides evidence for policy makers about the performance* *of public services, based on an assumption that such evidence leads to better policy. It is operated by a group appointed to do so, is owned by the government and operates within legal constraints using available finance and human resources.*

It is, of course, possible to conceive of a different form of performance measurement to see what works; one that is run by the service providers themselves in order to learn what works, enabling them to improve their performance. This becomes performance measurement to support learning and might have a root definition as follows.

Customers: in this case the immediate beneficiaries are the people who provide and manage the services in question, rather than the policy makers.

Actors: these are likely to be the staff of the service provider, aided by specialist analysts if that is appropriate.

Transformation: such a performance measurement system is intended to provide evidence of service quality to those who run the service.

Weltanschauung: this type of performance measurement rests on an assumption that collecting, analysing and feeding back information about performance, by those who provide that service, will enable learning that should lead to improved performance.

Ownership: without the active co-operation of the service providers such a system will not succeed, however it may still be within the remit of the government to close it down.

Environmental constraints: this too operates within legal constraints, plus those defined by the finance and expertise available, controlled by the government.

As before, this can be captured in a couple of sentences, defining such performance measurement as: A *system that provides feedback to service providers, operated by or on behalf of those providers, about the quality of service provision. It assumes that such feedback will support learning and continuous improvement and must operate with legal constraints using available financial and human resources and with the co-operation of the service providers and of the government.*

To identify competences

This form of performance measurement, unlike that conducted to see what works, has only a single general form to be captured in a root definitions. This measurement to identify competences is intended for the benefit of the funding bodies – central government in this case. Its CATWOE might look as follows.

Customers: since the performance evaluation is conducted to inform the funders, central government, then it seems reasonable to regard them as the main customers. As with the previous root definitions, It is of course true that the public should benefit from improved services, however the immediate beneficiaries are the central government.

Actors: were the proposals of the RSS, summarised in Bird et al (op cit) to be implemented, then an independent cadre of performance analysts would occupy this role (as with measurement to see what works). However, real life is somewhat different and so the best than can be said is that performance measurement to see what works will be run by a group appointed to do so.

Transformation: the aim of such a performance measurement system is to produce comparative performance indicators so that best practice can be identified. This must, of course, be done in the light of the purpose of the public service under scrutiny.

Weltanschauung: a worldview within which the rest of the root definition makes sense is one which states learning can occur across provider units if best practice and, poor performance can be identified.

Ownership: there seem to be two communities that can modify or close down such performance measurement systems and these are the political and policy communities as reflected in the government of the day.

Environmental constraints: the obvious ones to consider relate to the legal system, available finances and available skills, which will all directly affect such performance measurement in practice.

Hence, summarising this root definition in a couple of sentences leads to something like the following: *A system that provides evidence for policy makers about the relative performance of different providers of public services, based on an assumption that this will allow identification of best practice and of poor performance. It is operated by a group appointed to do so, is owned by the government and operates within legal constraints using available finance and human resources.*

To support public accountability

The previous two reasons for performance measurement in the delivery of public services have been relatively internal – internal, that is, to government, policy makers and service delivery organisations. However, performance measurement to support public accountability must, if it is to be done properly, have an external focus. Hence a suitable root definition might be as follows.

Customers: since the performance evaluation is conducted to inform the public, it seems reasonable to regard them as the main customers. It is of course true that the image of the government should improve if public services are seen to improve but this is a subsidiary concern in this case.

Actors: as with the previous root definitions, were the proposals of the RSS, summarised in Bird et al (op cit) to be implemented, then an independent cadre of performance analysts would occupy this role. However, real life is somewhat different and so the best than can be said is that performance measurement to see what works will be run by a group appointed to do so.

Transformation: the aim of such a performance measurement system is to produce information for public consumption that is an accurate and understandable indicator of the performance of public services. This must, of course, be done in the light of the purpose of the public service under scrutiny and with public priorities in mind.

Weltanschauung: given the customers, actors and transformation, it is clear that the worldview underpinning this root definition is that the public have a right to know how well public services are performing. Whether they are interested in this information is another question altogether.

Ownership: to the usual two communities that can modify or close down such performance measurement systems - the political and policy communities as reflected in the government of the day – we must also add the public itself.

Environmental constraints: the obvious ones to consider relate to the legal system, available finances and available skills, which will all directly affect such performance measurement in practice. We should also note that the relative interest of the public could be a constraint.

Summarising this root definition in a couple of sentences leads to something like the following: *A system that provides evidence for the public about the performance of public* *services, based on an assumption that the public has a right to know. It is operated by a group appointed to do so, is owned by the government and operates within legal constraints using available finance and human resources and should take account of the public's interest.* One problem with this type of root definition is that it assumes the existence of the 'public' and this turns out to be a far from straightforward assumption. This is amply demonstrated by Contandriopoulos, Denis and Langley (2004) in its demonstration of the problematic nature of the 'public' in Canadian health care.

To enable central control

Though there are clear linkages between the four different perspectives on performance measurement, though are also major differences, as should be clear from the root definition developed in this section. As before, the 6 elements of CATWOE will be used to develop the root definitions.

Customers: since the performance evaluation is conducted to ensure that public services are provided according to central plans and requirements, it is clear that the direct beneficiaries of the evaluation are the government and policy staff at the centre. They would, of course, argue, that this control is exercised in the public interest, but in root definition terms, the government and policy staff are the clear customers.

Actors: there seem to be three main sets of actors if this central control is to be successfully achieved. The first are staff in the units that provide the service (a hospital, benefit office, or whatever) whose job is to collect data that is submitted to the centre. The second group is the people in the centre who receive this data, analyse it and draw conclusions about the performance of the units. The third group is the government and policy staff who take action, if needed, to encourage good performance, using whatever incentives are considered appropriate.

Transformation: the aim of such a performance measurement system is to allow a central group to exert control over service providers. Thus the transformation, were such a system not in place, is to go from no control to whatever control is required.

Weltanschauung: if it were thought that local service providers would automatically follow central guidelines and requirements, then such central control would be unnecessary. Thus, such central control seems to be based on a view that without it, service providers would not operate as required.

Ownership: there seem to be two communities that can modify or close down such performance measurement systems and these are the political and policy communities as reflected in the government of the day.

Environmental constraints: the obvious ones to consider relate to the legal system, available finances and available skills, which will all directly affect such performance measurement in practice.

Expressing this CATWOE in a couple of sentences might lead to something like the following: *A system that enables government and its policy staff to exercise control over service providers whose performance is monitored on behalf of a central evaluation unit and this is justified because, without it, service quality would not improve and might deteriorate. It must operate with legal constraints using available financial and human resources and with the co-operation of the government.*

COMPARING THE ROOT DEFINITIONS

Root definitions, as developed in SSM by Checkland and colleagues, were intended to enable the comparison of the current situation by supporting a debate about alternative designs. That this debate should include consideration of different worldviews is crucial. It is important, though, to realise that any models developed from the root definitions are not models of the real world but, instead, are conceptualisations of aspects that would be expected in any attempts at implementation. In this paper, the root definitions are not being used for comparison with an as-is situation, but to support a debate about possible and desirable uses for performance measurement. Hence, the comparison is between the root definitions themselves. The previous section developed five root definitions, which are summarised in Table 1.

Similarities

It is easiest to start with the obvious similarities between the root definitions. It is clear from Table 1 that these are to be found in the O (ownership) and E (environmental constraints) of the CATWOES. Since the different systems are funded through central taxation, it should be no surprise that government itself and its advisors are regarded as the owners, as they have the power to shut things down. It might be argued that the real situation is more subtle than this, since it would not actually be possibly to operate such systems without some degree of acceptance by service providers and the general public. However, the direct power seems to rest with government and its advisors. This seems uncontroversial and pretty obvious.

	To see what works		**To establish competences**	**To support public accountability**	**To enable central control**
	External	**Internal**			
C	policy makers	service providers	government	the public	government and policy staff
A	specialists	service providers	specialists	specialists	service providers staff, analysts and policy staff
T	provide evidence for policy development	provide evidence of service quality	provide data to allow comparison of performance	provide reliable data for the public	maintain control
W	good policy is based on sound evidence	evidence of service quality allows improvement	need to do this to support learning	public has a right to know	important to maintain control
O	government & policy staff	government & policy staff	government & policy staff	government & policy staff	government & policy staff
E	law, finance & skills	law, finance & skills	law, finance & skills	law, finance, skills & public interest	law, finance & skills

Table 1: The root definitions summarised

Three of the root definitions assume that a group of measurement specialists will be the main actors in the measurement system. These are the root definitions for measurement as: the external attempt to see what works, the attempt to establish competences and the view that measurement is needed to support public accountability. They have in common a notion that measurement is externally imposed and that those providing the public service either should not be burdened with the measurement task or cannot be relied upon to do it properly. As commented earlier, the RSS Working Party (Bird et al, op cit) seem to assume that such a role could be occupied by its own members or by some apolitical body accredited for this purpose.

Differences

As we consider the differences, things start to get interesting. First, consider the different views of who will benefit from the measurement: in SSM terms, who are the customers of the system? If measurement is externally imposed to see what works, then it is clear that it is intended for the immediate benefit of those who develop and are responsible for public policy. Of course the measurement may be of interest to the public and of some value to the service providers, but the clear intention is that it should be used in policy development. It is hard to argue that this use of measurement is a bad thing, but there can be negative consequences, as will be discussed later. By contrast, if the measurement is internally driven by the service providers themselves so that they can see what works, then this is clearly for their benefit since it can support their learning and improvement. In the case of internally driven measurement, the performance indicators are probably not intended for external consumption, possibly from a fear of misinterpretation, but also from a concern that they might lead to external praise or blame, which will not support innovation and may reduce the learning and improvement.

Who are the customers if the aim of the measurement is to establish competencies? It seems that they are the people who hold the purse strings: central government in the case of national public services. As with the external attempt to see what works, there are obviously other groups interested, but the main beneficiaries are those who foot the bills, since it enables them to identify the good and poor service providers. It could be argued that the next root definition, measurement for public accountability, is similar: the public rather than the political classes can see where standards are high. If we believe that people have a right to choose between different providers of public services, then this form of measurement and its publication are vital. Finally, what of measurement to enable central control? This is an extreme form of measurement to establish competences, with the added rider that the customers (government and policy staff in this case) will act to reward good performance and to punish poor performance: praise and blame. If any learning occurs in the provider units, it is based on a stimulus:response model which may lead top highly defensive behaviour – hardly the best way to encourage innovation and creativity.

It is, of course, in the essential transformation and Weltanschauugen that the major differences in the five root definitions are most obvious. In the two cases of measurement to see what works, the essential transformation (what the system does) is the provision of evidence. In one case this is for policy development in the belief that evidence-based policy is better than that plucked from the thin air of political whim. In the other case, this is to help the service providers understand the good and bad points of their operations in the belief, as mentioned above, that this will support their learning. If the services provided have some degree of repetition and if outcomes can be related to actions taken, this form of measurement goes some way to meet the criteria set by Hofstede (1981) for improvement by routine control. Moving to the right of Table 1, the next column represents measurement by agents of the government to establish competences, with its implicit notion of comparison. As with internal measurement to see what works, the justification for this measurement is that it will support learning. However, it should be noted that it is the central government or its agents that learns, rather than the service providers. If the two approaches to measurement could be aligned without coercion, then it may be possible to have the best of both worlds.

The next column of Table 1 represents measurement to support public accountability and hence its essential transformation is to provide information for the public, based on the straightforward premise that the public has a right to know. As mentioned earlier, it turns out to be harder to define what we mean by the public that might be the case. For example, Contandriopoulos, Denis and Langley (op cit) discuss this in the context of Canadian healthcare and identify three different 'publics' whose support is often solicited by politicians. The first group they call the *reified public* and it includes the 'real public', the 'average man or woman' and the 'ordinary citizen'. This is an abstract group with no involvement in the healthcare system and has "the same status as the abstract citizen upon whom the democratic system rests." (p1580, op cit). This reified public occupies a privileged place since its members are thought of as 'average men and women' who are, almost by definition, relatively ignorant and in need of education and information. Anyone who gets involved in the system will lose their status as 'average man or woman'. They call the second group *the public's voices* and it includes those who participate in decision-making processes such as citizens' representatives, community organisations, citizens' groups etc.. These people project themselves as effective public representatives, but others in the system often do not see them in these terms, rather seeing them as partial and with conflicts of interest. They are engaged in a symbolic struggle for the right to speak for the 'public' and may attempt to present themselves as 'average men or women'. Finally, there are the users, patients, families and others who consume services and who have opinions and interests that differ significantly from those of the 'general public'. Which group is the public who will act as customers for this fourth root definition? Which is the public that has a right to know and what do they have a right to know? How can performance data be provided that is meaningful and is it to be aimed at the uninformed reified public, the public's voices or at service users?

Finally, table 1 shows measurement to enable central control, which pretty much defines the transformation of the root definition – to enable those who hold the purse strings to remain in control. This is based on a belief that any improvement in service quality will only be achieved if central control is maintained. Thus information is collected at the behest of the centre and is used there to engineer whatever changes it feels are necessary.

A call for clarity

The preceding discussion covers five different justifications for the use of performance measurement in the delivery of public services. It is obviously possible to design performance measurement systems that fit more than a single column of Table 1, but this will not happen by accident. It will only happen if there is an explicit debate about the intended purpose of the measurement system. It is possible to make a sensible case for each of the five root definitions and this is evident in the fact that their underpinning Weltanschauungen do make sense as worldviews that are appropriate justifications. However, it is pointless to pretend that organisational learning and innovation is encouraged by externally imposed regimes that feature praise and blame. Thus the argument presented here can be summarised as follows:

1. There are four obvious reasons for the use of performance measurement in the delivery of public services: to see what works, to identify competences, to support public accountability and to allow control of decentralised service provision.
2. Any attempts to design and implement such performance measurement systems should include explicit consideration of their underpinning worldviews and the transformations that follows from this. A system can meet multiple objectives, but is much more likely to do so if designed for this purpose.
3. It is highly unlikely that an externally imposed measurement system that aims to enable central control will also be useful in supporting innovation, experimentation and learning.
4. If the performance indicators are intended for public consumption, it is important to be clear who this public is and it may be necessary to use expert intermediaries to explain the significance of the observed performance.

ACKNOWLEDGEMENT

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

REFERENCES

Ackoff R.L. (1974) Redesigning the future: a systems approach to societal planning. John Wiley and Sons, New York.

Bird S.M., Cox D., Goldstein H., Holt T., Smith P.C. (2003) Performance indicators: good, bad, and ugly. Report of the Royal Statistical Society Working Party on performance monitoring in the public services. Royal Statistical Society, London.

Boyne G.A., Farrell C, Law J, Powell M. and Walker R. (2002) Evaluating public management reforms: principles and practice. Open University Press, Milton Keynes.

Brown J, and Cooper C. (2004) The complementary use of hard and soft or in developing tax policy. In Pidd M. (2004) (ed) Systems modelling: theory and practice. John Wiley, Chichester.

Brown J., Cooper C. and Pidd M. (2005) A taxing problem: The complementary use of hard and soft OR in the public sector. In press.

Charih, M. and Rouillard, L. (1997) The New Public Management, in Charih M. and Daniels A. (eds) New Public Management and Public Administration in Canada, pp. 27–45. Toronto: Institute of Public Administration of Canada and École nationale d'administration publique.

Checkland P. and Holwell S. (1998a) Information, systems and information systems, John Wiley, Chichester.

Checkland P.B. (1981) Systems thinking, systems practice. John Wiley, Chichester.

Checkland P.B. (2001). Soft systems methodology. In Rosenhead J.V. and Mingers J. (2001) Rational analysis for a problematic world revisited: problem structuring methods for complexity, uncertainty and conflict. John Wiley, Chichester.

Checkland P.B. and Scholes J. (1990) Soft systems methodology in action. John Wiley, Chichester.

Contandriopoulos, D, Denis, J-L and Langley, A (2004) Defining the 'public' in a public healthcare system. Human Relations, 57(12): 1573–1596.

Goldstein H. and Spiegelhalter DJ. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance (with Discussion). Journal of the Royal Statistical Society Series A, 159: 385 – 443.

Gow, J. I. & Dufour, C. (2000) Is the New Public Management a Paradigm? Does it Matter? International Review of Administrative Sciences 66 (4) : 573-598.

HM Treasury (2004) http://www.hm-treasury.gov.uk/media//CF0A5/sr04_psa_ch6.pdf

Hofstede G. (1981) Management control of public and not for profit activities. Acctg, Orgs and Soc, 6, 3, 193-211.

Kaplan R.S. and Norton D.P. (1996) The balanced scorecard: translating strategy into action. Harvard Business School Press, Boston, Mass.

Manning N. (2000) The New Public Management & its legacy. http://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. JAMA.283:1866-1874.

Miller D. (2003) The virtual moment. Royal Anthropological Institute 9 (1), 57-75.

Mingers J. and Taylor S. (1992) The use of soft systems methodology in practice. Journal of the Operational Research Society, 43, 4, 321 32.

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.

Pidd M. (2003) Tools for thinking: modelling in management science (2nd Ed). John Wiley, Chichester.

Pidd M. (2005) Understanding perverse effects of public sector performance measurement. In Hartley J. and Skelcher C. (2005) Managing improvement in public service delivery, OUP, in press.

Smith P.C. (1995) On the unintended consequences of publishing performance data in the public sector. Intl Jnl Pub Adm, 18, 2 & 3, 277-310.

Stationery Office (2003) On target? Government by measurement. HMSO, London.

Wilson B. (2001) Soft systems methodology : conceptual model building and its contributions. John Wiley, Chichester.