

Occasional paper no. 21

Performance measurement: Foundations, consequences and futures

Professor Jenny M Lewis

state**services**authority



The Australia and New Zealand School of Government and the State Services Authority of Victoria are collaborating on a partnership to build connections between new thinking, practice and implementation in public administration. The Occasional Paper series is part of the partnership program.

About the Australia and New Zealand School of Government

ANZSOG is a consortium of governments, universities and business schools from Australia and New Zealand. It represents a bold commitment on behalf of member governments to strengthen the management and policy capacity of their respective public sectors.

ANZSOG is a world-class centre that provides tailored learning opportunities for future leaders of the public sector. ANZSOG's purpose is to encourage improved public sector decision making, leadership and policy outcomes for the benefit of the whole society. In doing so, the School also plays a key role in promoting public service as a profession of great social value.

ANZSOG has three core activities:

- executive education courses including an Executive Master of Public Administration degree;
- a case study program; and
- a research program

These programs aim to deepen knowledge and understanding of government and to disseminate that understanding to the wider community – for the benefit of the whole society.

Contact us at:

ANZSOG
email: anzsog@anzsog.edu.au
phone: +61 3 8344 1990
fax: +61 3 9349 5849

Postal Address:

PO Box 230
Carlton South
Victoria 3052
Australia

Web address:

www.anzsog.edu.au

About the State Services Authority

The State Government of Victoria has vested the State Services Authority with functions designed to foster the development of an efficient, integrated and responsive public sector which is highly ethical, accountable and professional in the ways it delivers services to the Victorian community.

The key functions of the Authority are to:

- identify opportunities to improve the delivery and integration of government services and report on service delivery outcomes and standards;
- promote high standards of integrity and conduct in the public sector;
- strengthen the professionalism and adaptability of the public sector; and
- promote high standards of governance, accountability and performance for public entities.

The Authority seeks to achieve its charter by working closely and collaboratively with Victorian public sector departments and agencies.

Contact us at:

State Services Authority
Email: info@ssa.vic.gov.au
Phone: +61 3 9651 1321
Fax: +61 3 9651 0747

Postal Address:

3 Treasury Place
Melbourne
Victoria 3002
Australia

Web address:

www.ssa.vic.gov.au

Published: July 2012

Series foreword

The Australia and New Zealand School of Government and the State Services Authority are collaborating on a partnership that draws together a broad network of policy-makers, practitioners and leading academics.

The partnership is designed to build connections between new thinking, research and practice in public policy and public administration.

The Occasional Papers explore the challenges and opportunities in public administration. They showcase new ideas and offer new insights into issues facing the public sector.

Written by either academics or public servants, the papers bring together the academy with public policy practitioners.

We trust that you find the Occasional Papers stimulating and thought provoking. All papers in the series are published on the ANZSOG and SSA websites.

Professor Allan Fels AO
Dean
Australia and New Zealand School of Government

Bruce C Hartnett
Chair
State Services Authority

About the author

Professor Jenny Lewis is currently a Professor in the Department of Society and Globalisation, Roskilde University, having formerly been Associate Professor in Public Policy in the School of Social and Political Sciences at the University of Melbourne. Jenny has worked for the Victorian State Government in a number of different roles and as a consultant. Her research is centred on new forms of governance, the public policy process, performance measurement, and professions and new public management. She has recently been awarded a prestigious ARC Future Fellowship to examine performance measurement in the health and higher education sectors in four countries – Britain, Australia, the Netherlands and Denmark.

Performance measurement: Foundations, consequences and futures

Author

Professor Jenny M Lewis
Department of Society and Globalisation
Roskilde University, Denmark

Performance measurement is necessary, and it is not going away. It is never far from the minds of senior managers in public sector organisations. After the last 30 years, during which time its importance has grown, it is time to take stock of how performance measurement has been understood and how its impacts have been analysed. This paper examines why the public sector is interested in performance measurement, how it came to be more important, and outlines some of the fundamental assumptions underpinning it. It considers the consequences of performance measurement, and discusses a recent shift in thinking that focuses on its potential to become a more positive force for change.

In summary, this paper uses insights about unanticipated consequences, constant error correcting, and the chaining of social problems to create a performance measurement chain. It argues that this will assist practitioners to consider ways in which the undesirable consequences of performance measurement can be minimised.

Why performance measurement?

On paper, the fundamental purpose behind performance measurement is to improve performance. Beyond that deceptively simple starting point, there are many different reasons why it might be done. It may be for the purposes of evaluation, to discover if a program is doing what it is supposed to do. It might be used in order to control the performance of those working in a program or service of interest. It might also be simply a means of controlling the budget for a particular area, or determining how much money is being spent to achieve some desired output. There are many motivational issues at stake, because performance measures aim to discover what is working and what is not – or more personally, who is doing what they should be and who is not.

Linking measurement to management, it becomes clear that an important reason for measuring performance is that work that is not measured cannot be managed. If there are no criteria for measuring the value of what is being produced, then there is no way of arguing that it is better than something else that could be done instead. Performance measurement is important as a means to discover what needs to be improved, which individuals and organisations are meeting their goals, and which are failing to meet the established targets and need either help or penalties. In short, there are many good reasons for wanting to measure performance, and the pressure to do so has only increased over the last three decades.

New Public Management and performance models

Why has performance measurement apparently become more important for governments and public sector organisations? A substantial part of the answer is the rise of New Public Management (NPM). With its basis in both scientific management ideas and transaction cost economics, NPM was meant to be a politically neutral, rational means for governing. And indeed, it has been embraced by both conservative and progressive governments, as they have become increasingly focused on saving money, reducing the time and effort expended, and cutting waste. In Hood's (1991) famous description

of NPM, he argues that it was frugality and the reduction of waste that became its single minded focus. This is not in itself bad – it is important to account for a wise use of public funds, but it is also prudent to note that other desirable attributes (justice, equity) may be downplayed or even ignored if this becomes the singular point of concern.

Performance measurement was (and is) a central plank of NPM reforms, with the link from inputs to outputs a major concern. As NPM gained ascendancy, a newfound belief that previous problems could be avoided if there were more measurement and more management became widespread. It was assumed that if the right systems could be put into place, then it would be a relatively easy task to identify efficiency savings while also continuing to deliver the desired outcomes.

When deciding upon what constitutes performance, and how it should be measured, a basic production model is generally used. This follows an engineering approach where inputs are transformed into outputs. Using this as a foundation, a performance model can be represented quite simply, flowing from inputs, through processes and outputs, to outcomes (see Figure 1). To make this model more aligned with the concerns of the public sector, rather than the production of widgets, a set of directions from stakeholders flowing into policy making and objective setting can be added.

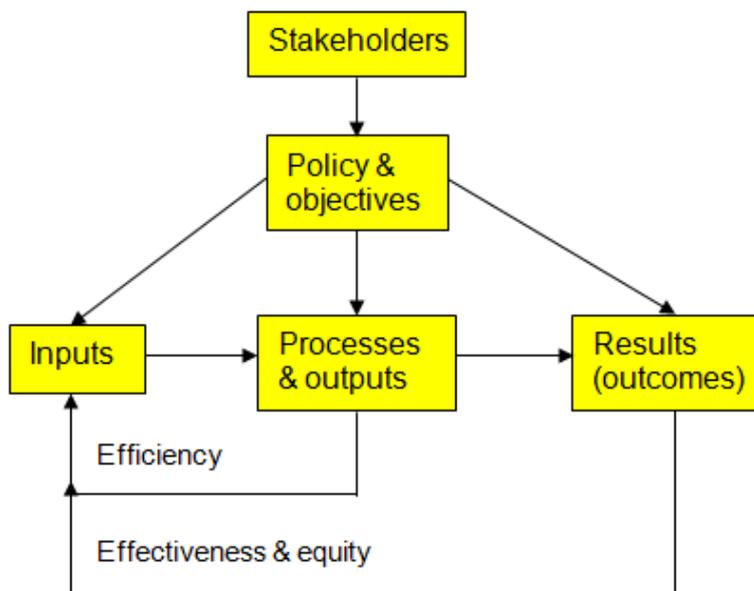


Figure 1: A simple performance model (from Johnsen 2005)

The three boxes from inputs to results, shown in this fairly typical performance model, are used in calculating measures of performance. Efficiency is defined as the ratio of processes and outputs to the inputs. Effectiveness is the relationship between outcomes and inputs, and equity is the distribution of outcomes across different individuals and groups in relation to inputs.

These models, of course, ignore such things as the uncertainty of policy and objectives, and the often widely differing stakeholder views. Even if such a model could factor in important contextual considerations, there are other measures of performance that might be considered important, and the same measures can have different and also more complicated definitions.

For example, economy (rather than efficiency or effectiveness) might be the main priority. If so, the focus might simply be on measuring total costs, or actual versus expected costs, or on minimising the

costs of a particular program. Efficiency, rather than being the ratio of inputs to outputs as defined above, might be interpreted as meaning achieving the output for the least cost, or achieving a specified number of outputs for the least cost. Effectiveness may be measured in terms of achieving the program's intentions, or the extent to which the output achieved the stated objectives, or whether a particular set of targets are achieved. Equity, as a measure of the distribution of outcomes, might signal that measures will demonstrate equity if all similar cases are dealt with in like fashion. Alternatively, an equity measure could focus on whether a program is equitable with regard to the process, or in terms of the outcomes. And finally, quality, which is notoriously tricky to define and quantify in many programs, could be the real measure of interest.

A further consideration is whether and how the information collected is actually used, by whom, and for what purposes. It is interesting that the simple performance model shown in Figure 1 does not include any feedback loops to policy making and objective setting. A comparative study of several European countries found that Finland, the Netherlands, Sweden and the UK are all now doing more performance measurement (Pollitt 2006). In each country, these performance measures are most often used by managers, not by politicians, and not by people further down the hierarchy. It makes sense, since the rise of measurement occurred alongside the introduction of NPM, that the people who are charged with managing performance utilise the measures most.

Founding assumptions of performance measurement

As this very brief discussion highlights, performance models can involve a variety of measures, and can be used in diverse contexts. But all performance measures involve some vital assumptions, the first of which is that it is possible to actually measure the outputs or outcomes of interest. In other words, it is assumed that the indicators used do actually represent, at least with some degree of validity, the result of interest (Bevan and Hood 2006).

While it is accepted that it is impossible to capture everything that represents the 'good' of interest, there is an assumption that the core of the 'good' can be measured with some accuracy, but this does not always hold. For example, test results are seen to be an indicator of how a group of students are progressing in their studies. Bad results might indicate an unmotivated group of students, a poor standard of teaching, or a lack of resources, but the results themselves do not disclose which of these (if any) is occurring.

The second assumption is that 'gaming' of any measurement system will occur, but it will be relatively small (Bevan and Hood 2006). Measurement is assumed to change behaviour, and there will be some gaming in any system, but this is considered to be a relatively minor component of the behaviour change that the system will set in train. This turns out to be a rather large assumption in practice. While the term 'gaming' suggests consciously bad behaviour, this is not always the case. It may be a very deliberate choice, but it may simply be a means for meeting the required targets in difficult circumstances.

As soon as performance measures are put into place, people begin to change what they do. If individuals are aware that the data collected will actually be used in some way that affects their work, particularly in controlling *how* they work, then they will surely react. It makes little difference whether the measures are consciously integrated or not, and whether individuals agree with them or not. Everyone will be measured against the standards, regardless of their personal knowledge and acceptance of the measures.

Some of the things that induce gaming are easy to observe, if the emphasis is on understanding how people react to targets (Bevan and Hood 2006):

- 1) The ratchet effect: people in positions that they intend to still be occupying in 12 months' time are reluctant to exceed any targets (at least by too much) if it seems likely that this will lead to the target being ratcheted up next time. In other words, people have a disincentive to perform above the target set.
- 2) The threshold effect: if people are already exceeding the target then they have no incentive to keep over-performing. The incentive is to work at the level of the target or a level just above it.
- 3) Output distortion: this occurs when people only do what is measured. If there is something that is more effective, but it does not count in the measures, then it will not be done, or it will only be tried if there is time to do it after the measured things are completed.

The consequences of performance measurement

The central concern of this paper is the consequences of introducing performance measures. If performance measurement is important for governments and managers, what kinds of impacts – both desirable and undesirable – does it have?

This topic has not been widely studied, which is surprising given the importance of sociological and psychological aspects to understanding the effects of performance measurement. How people react to being measured has substantial consequences for the individuals themselves, but also for performance measurement systems. Three scholars provide helpful insights – although none of them specifically wrote about performance measurement. They were each concerned with social problems, and they provide some significant insights that can be usefully applied to this topic. They are:

- Robert Merton (1936) and unanticipated consequences
- Charles Tilly (1996) and the invisible elbow
- Gary Fine (2006) and the chaining of social problems

Unanticipated consequences

Merton (1936) described how purposive actions (actions that involve motives and consequently a choice between various alternatives) can fail to produce the expected outcomes. The consequences of actions result from an interplay of the action and the situation. Purposive action involves motives and choice, but not always clear aims (and not necessarily rationality of aims). There are problems, then, of causality and knowing what the purpose is. There are limitations to what can be anticipated to result from actions. A lack of adequate knowledge is one limitation, and deviations can be anticipated, but not fully. Opinion and estimate play an important part in deciding on a course of action, but there is always potential for important considerations to be disregarded. Time and energy are also limited, rendering it financially and psychologically impossible to take into account all the relevant factors.

Errors are another important source of unanticipated consequences: incorrect appraisals of a situation are common, as are incorrect predictions, the selection of incorrect actions, and the faulty execution of a chosen action. Humans rely on habits (whatever worked before), and neglect important elements of the problem. There is a tendency to satisfy immediate interests and neglect longer term consequences, and, of course, choices are not unbiased but are based on a specific set of values. Activities set off processes that have far reaching changes (which are not limited to the specific area intended), and predictions themselves can become a new element in the situation, changing the set of things that have impact (Merton 1936).

In relation to the topic of this paper, it is evident that when a performance measurement system is put into place, some of the resulting consequences will be things that were expected to happen, but many of them will not. The unintended (and undesirable) consequences of measurement include such things as cheating, bribery, 'teaching to the test', over and under reporting in order to look good, incentives to not disclose problems, creaming of the easy to achieve targets (and neglect of the hard cases), an inflation of quantity and a lowering of quality, and a raft of false economies (Grizzle 2002).

The invisible elbow

Moving on to incessant error correction, it is useful to remember that people instantly recognise the effects of what they are doing, and they constantly adjust to the situation as it unfolds, in the face of things that were not expected to happen. Tilly (1996) describes how all actions lead to some recognition of more problems, which people deal with through a series of improvised adjustments. He suggests that Merton told only half the story – making the case for why purposive action regularly generates unanticipated consequences, but not going on to suggest that this action produces social structures. Tilly (1996) argues that the consequences of all social interaction pass through powerful systematic constraints, which are created by the impact of cultures and relationships. There is, then, a great need for the examination of errors, their consequences, and their rectification – what he refers to as incessant error correction (or the *invisible elbow*: when we do things with our hands, as beneficial as they may be, we often inadvertently disturb other things with our elbows).

The chaining of social problems

Fine (2006) similarly notes a kind of path dependency of social problems, where consequences lead to more problems and new solutions being put into place. This generates a chain of social problems, sets the term for later problems, and constrains the options based on earlier decisions. The consequences of action result from an interplay of the action and the situation. And in line with Tilly, he notes the constant need to adjust and correct small mistakes, which are all based on the accumulation of what has happened before (culture) and social relations and interactions.

These insights about the unanticipated consequences of purposive action; the constant error correcting that occurs as events play out; and the chaining of social problems, generating a social system of opportunities and constraints, can be applied to performance measurement. To understand the consequences of performance measurement, it is necessary to see it as a chain – a set of interactions that have a set of consequences, which leads to a set of corrections of the unanticipated and undesirable effects, which generates a social structure of performance measurement.

The following, more detailed performance measurement chain should be examined, in order to understand its consequences:

- *Context* – social, economic, and political institutions shape the context within which policies are made.
- *Intentions* – decisions are made about what is valued based on what is intended.
- *Criteria* – measurement criteria that attempt to capture these values are established.
- *Sent rules* – system rules are made that try to measure what is valued, using these criteria.
- *Received rules* – those who are measured receive instructions about performance measurement.
- *Understandings* – those who are measured come to have (not necessarily accurate) knowledge of what is being measured and why from a variety of sources.
- *Actions* – they respond to the signals received based on those understandings.
- *Consequences* – their actions have a series of consequences, both intended and unintended, and both desirable and undesirable.

Approaches to measurement and overload

Over the last decade, some movements away from the *managerial* way of thinking about performance measurement (focused on punishment), towards what is known as a *learning approach* (which uses terms such as celebrating, promoting and learning) have been noted in the performance measurement literature (Behn 2003). This altered focus generates a less punitive style of performance measurement: rather than concentrating efforts on publicly shaming those who have failed to meet the targets set, such systems should be geared towards finding the high achievers, giving them credit, and ensuring that learning occurs and good practice is spread (Bovens et al 2008).

While this move rests upon an inherently positive notion, it has the potential to overload measurement systems. Since an important concern is whether the information provided by performance measures justifies their enormous cost (Jackson 2011), a consideration of this move and its consequences is an essential consideration here. A brief contrast of the components of these two different approaches is summarised in Table 1.

Table 1: Two approaches to performance measurement

	Managerial	Learning
<i>Creation of measures</i>	Hierarchical – imposed on those measured	Interactive – co-created with those measured
<i>Level of control</i>	Central	Local
<i>Key driver</i>	Market / competition	Collaboration
<i>Key concern</i>	Outputs	Processes
<i>Measurement focus</i>	Failure	Quality
<i>Measurement scope</i>	Crude and simple	Sophisticated and comprehensive

Source: adapted from Lewis and Triantafillou (2012)

In short, managerial systems view performance measurement as relatively simple and quick, while a learning approach makes it more complicated and time consuming. The question then becomes, if better measures are created and implemented, will that improve the situation? While notionally the answer to that question is a fairly straightforward affirmative (since such a move appears to be built on positive intentions), there are some cautions that should be expressed.

First, a learning approach is more likely to supplement than replace existing measures, as the simple indicators will almost certainly remain useful. Second, it is expected to require more rather than less data, as the indicators are more sophisticated. Third, it will probably increase compliance costs, because it requires more participation and dialogue with those who are measured. Fourth, it has the potential to create a ‘change for the sake of change’ mentality – if you are not learning then you are going backwards, and the pressure is likely to fall on those who are not seen to be trying to innovate, even if there is no reason to (or bad reasons to).

This raises the question: How much measurement is really essential if overload is to be avoided? The following guidelines are suggested. In defining performance measures, there is a need to:

- delimit the scope of measures – avoid “some is good, more is better”;
- examine the costs and benefits of measuring and remove any unnecessary measurement tasks;

-
- consult those who are measured to use their experience to help improve the effectiveness of implementation; and
 - analyse not just what performance measurement is, but what effects it has on a range of actors, at a whole series of junctures.

Returning to where this paper started, it is time to reiterate that performance measurement is here to stay and it is necessary. To improve how performance is measured more generally, there is an urgent need to understand and examine all the links in the performance measurement chain outlined in this paper, and the many consequences associated with each of them. Only if the whole chain is considered can better methods for minimising the undesirable consequences of performance measurement be discovered.

References

- Behn (2003) 'Why measure performance? Different purposes require different measures', *Public Administration Review*, 63(5): 586-606.
- Bevan, G and Hood, C (2006) 'What's measured is what matters: Targets and gaming in the English public health care system', *Public Administration*, 84(3): 517-538.
- Bovens, M, Schillemans, T and 't Hart, P (2008) 'Does public accountability work? An assessment tool', *Public Administration*, 86(1): 225-242.
- Carter, N, Klein, R and Day, P (1992) *How organisations measure success: The use of performance indicators in government*, London and New York: Routledge.
- Fine, G (2006) 'The chaining of social problems: Solutions and unintended consequences in the age of betrayal', *Social Problems*, 53(1): 3-17.
- Grizzle, G (2002) 'Performance Measurement and Dysfunction: The Dark Side of Quantifying Work', *Public Performance & Management Review*, 25(4): 363-369.
- Hood, C (1991) 'A public management for all seasons', *Public Administration*, 69: 3-19.
- Jackson, P (2011) 'Governance by numbers: what have we learned over the past 30 years?' *Public Money & Management*, 31(1): 13-26.
- Johnsen, Å (2005) 'What does 25 years of experience tell us about the state of performance measurement in public policy and management?' *Public Money and Management*, 25(1): 9-17.
- Lewis, J and Triantafillou, P (2012) 'Accountability as learning: A new source of government overload?' *International Review of Administrative Sciences*, 78(4), to be released in December.
- Merton, R (1936) 'The unanticipated consequences of purposive social action', *American Sociological Review*, 1: 894-904.
- Pollitt, C (2006) 'Performance Management in Practice: A Comparative Study of Executive Agencies', *Journal of Public Administration Research and Theory*, 16(1): 25-44.
- Tilly, C (1996) 'Invisible Elbow', *Sociological Forum*, 11(4): 589-601.