

Critiquing Performance Management Regimes and their Links to Budgetary Decision-making and Resource Allocation

John Wanna

Australian National University

Arguments

My main motivation for writing this article is to dissuade governments and their public service advisers from the simplistic adoption of flawed performance management regimes, in the mistaken belief that these systems will improve both delivery performance and effective resource allocation. While performance management offers attractions to governments and legislatures (and has many willing and enthusiastic adherents) there are inherent distortionary dangers in adopting and using flawed systems.

The main arguments of this article directly challenge: (i) the contestable assumptions underlying performance monitoring in public sector environments, (ii) the deplorable implementation of performance measurement in practice; and (iii) the unfortunate outcomes of performance management in government.

In relation to the underlying assumptions, I argue we invest too much blind faith in the benefits and reliability of performance monitoring; and that we assume and often take for granted that certain selected indicators can provide valuable assessments of the performance of public delivery actors.

In relation to implementation, I suggest that we do not yet have a good grip on measuring and assessing performance in the public sector. We have developed relatively poor indicators and measuring metrics and many of the performance indicators we have are often distortionary (and possibly perverse). Indicators also tend to capture the conspicuous phenomena and observable criteria not necessarily the substantive issues or measures of effectiveness. As a general rule, performance measuring regimes tend to be self-calibrating rather than tailored to the specific needs of programmes or their context/starting points. Moreover, those responsible for establishing performance monitoring regimes in the public sector still tend to prefer centralized systems of monitoring, seeking to standardize measuring requirements and reporting formats, and establishing generalizable monitoring systems when an allowance for greater diversity may better capture real levels of effectiveness.

In terms of the outcomes of performance management, I would argue we tend to have even poorer evaluation and review capacities when judging performance-related data. When we produce performance information it is generally the case that limited use

is actually made of such information by governments and relevant overseers (and almost none by the citizenry). And, when thinking of future developments we should be very wary about trying to link performance monitoring and reporting of activities too closely with subsequent resource allocation, budgetary decision-making and the alluring mirage of 'performance budgeting'.

Reasons for Measuring Performance

In recent decades, there has been a huge international interest in performance measurement and performance management across the Americas, Europe and Central Europe, Asia and the Pacific, and even parts of Africa (see Bouckaert and Halligan 2008 for an insightful international coverage). The high level of interest among governments represents a somewhat pious intellectual investment in the hope and aspiration that reliable evidence can be produced and used to benefit policy-making processes. But let us remind ourselves as to why we are measuring performance.

Performance in public sector agencies and programmes is not measured for its own sake. Rather assessments of performance are undertaken with some normative hope of improving certain policies, programmes, activities or operations, organizational achievements and community outcomes. It may be conducted by an external reviewer, professional evaluator or expert consultant, various committees, sometimes by programme managers and operational staff, or even by clients and citizens (end-users).

Performance monitoring and reporting can be undertaken for various reasons and at different stages of policy processes: it can be a rationale for policy interventions (initial motivation); it can be a deliberate aspect of good planning and design (formulation); it can be a form of comparative assessment (imposing competitive pressures or rankings); it can be a device for urging continuous improvement (productivity or effectiveness), or a defence and justification (a shield). However, in practice many examples of performance monitoring are patched-up exercises, very ad hoc and partial in nature and imposed ex-post well after the event or point of decision. Policy deliverers may have external indicators imposed upon them, sometimes relying on metrics or processes on which they were not particularly focused.

Yet, performance assessment (like coordination) is thus largely an assumed 'good', necessary beyond question or dispute (although there may be vehement disputes over the types and methods of information gathering). Many public management texts routinely present performance measurement as a vital component of good governance, an essential part of defensible policy processes, and a core responsibility of management. It is also seen as an important dimension of public accountability for service delivery and the deployment of resources, authority, equity or public value.

However, it should be noted that there is a strong tendency to centralize and standardize monitoring systems, to attempt to ensure comparability in reporting formats. But comparability can be a straitjacket. Performance measuring regimes in the public

sector tend to be run as bureaucratic exercises, becoming self-referential and self-calibrating rather than premised on the specific needs of programmes or their context/starting points. Performance regimes are thus expressions of power relations, potentially empowering some actors while disempowering others. Political executives and legislatures can impose performance regimes to 'control the reins', impose their particular priorities or extract savings. Executives may invest in performance monitoring as a way of protecting themselves and managing political risk. There can be a degree of distrust and cynicism shown towards performance measurement.

Performance reports are potentially utilised by different actors in the policy process. Ministers and politicians may view them as evidence of policy effectiveness or the net worth of additional resource investments. Yet they are risk averse wishing to avoid embarrassment and the full glare of scrutiny. Some managerial actors may use information for executive feedback on the progress and performance of policies/programmes. It is not necessarily a public process or transparent. Accountable actors (such as legislators, scrutiny bodies or auditors-general) may view performance data to gauge efficiencies and effectiveness to the extent they can. The media can use performance information in producing newsworthy stories, although only 'bad news' items tend to appear regularly. Unions or professional associations may use performance data to help regulate their employment practices, workloads or training requirements.

There have also been different waves of interest internationally in performance measurement. Many nineteenth century western governments were motivated by parsimony and imposed economies. Americans became interested in organizational performance ahead of most jurisdictions with the 'performance movement' of the early twentieth century feeding into broader reform efforts. Management was motivated by PPBS, MBO, ZBB, and later by performance based laws (GPRA, PART). This was accompanied by the development of extensive cultures of evaluation in post-war public policies—especially through randomized sampling and the highly quantitative use of management-derived surveys (open to agency bias) (see Radin 2006). Many other advanced nations began to focus on efficiency and technological improvements from the 1960s, with many developing nations following suit from the 1990s.

When governments embraced business-like practices from the 1980s onwards (managerialism, New Public Management, devolution, outsourcing) they often embraced performance measurement as a co-requisite to indicate improvement in management practices and outputs. Many governments announced extensive plans to evaluate programmes (the holy grail of policy evaluation and programme evaluation, which has often proved elusive, see Wanna et al. 2010). Over time, there has been a steady improvement in reporting arrangements and in their consistency, and generally many delivery areas of public policy now have pre-specified and timely reporting requirements; however such progress has not been linear (there has been some backsliding and gaming—especially between playing off input reporting versus output

reporting, and there remain questions over the quality of the information captured). Governments that dogmatically impose performance requirements on agencies can often generate resistance and push-back, and even non-compliance. Performance measurement even under more flexible business models of administration remains a 'work-in-progress'. In some jurisdictions arms-length monitoring bodies have also become more significant and can name and shame agencies publicly.

More recently, many governments internationally have committed themselves to realign public services through citizen engagement to better reflect citizens' preferences and satisfactions. This is ostensibly a means of reconceptualising performance from a client-oriented perspective. However, to date there is little evidence that citizens' expectations, citizen surveys, or satisfaction indices have been widely used in seeking feedback concerning the quality of services delivered by governments. There are some examples of nations undertaking regular citizen surveys of satisfaction (e.g. New Zealand, Canada) but not much close integration of this information with line delivery (see Lindquist et al. 2013 for some evidence of progress).

How do we Develop Performance Indicators?

Indicators of performance are often surrogates. We grab what we can; we collect the collectable; we measure the measurable. Indicators have many limitations. If indicators are merely surrogates they may be readily collectable but tell us very little of importance. Criteria we can count and measure tend to be measured, and therefore become manufactured by the producers. There may be no causal link between the indicator(s) collected and the desired outcomes expected. Agencies may simply 'tick the box' to comply with the collection of data but it makes no difference to the performance of the organization. Perverse logics also apply in that many agencies will be tempted to generate results only in those criteria measured which may entirely distort their behaviour and in some cases be counter-productive. For instance, measuring police performance by arrest rates is a simple enough metric, but it provides incentives to police to arrest citizens for any slight misdemeanours, to record all arrests or make multiple arrests, and to show little discretion in making arrests. It can have other unintended consequences such as filling the courts and gaols with relatively non-serious offenders. Police may be tempted to arrest and imprison fine-defaulters who are not really criminals. Hence, performance indicators are open to gaming and deliberate cheating—easy pass marks in educational exams, re-admissions in hospitals, focusing on 'easy' clients to service.

Hence, crude performance indicators are often restricted to easily identified and observable criteria, but may be limited in their actual applicability and usefulness. Many of such indicators only indicate levels of activity, or the 'busy-ness' of agencies/deliverers (they are volume measures). Other indicators can capture process data (such as the number of meetings held, cabinet submissions processed, inspections undertaken), which provides a measure of public service routines or throughput.

Quantifiable indicators can be presented as positive or desirable achievements, or as negative features to be lessened or avoided. Positive indicators include desired targets, stretch targets or number of processed items completed; workload estimates; programme outputs. Negative indicators include such measures as backlogs, queues and waiting lists, losses of clients or consumers (say declining educational enrolments). Some indicators express ratios of inputs to outputs, but as productivity measures these ratios are dependent on many other assumptions and calculations—e.g., what fixed and variable costs are included in the inputs (including overheads) and what data on outputs can be generated. For instance, the costs of running a public school may be itemisable, but what departmental overheads and central government costs should be allocated to the school? Moreover, apart from such measures as school attendance indicators, what outputs in terms of quality of education can be quantified to weigh against the consolidated input aggregates? And, if we are interested in productivity improvements are we interested in organizational productivity (internal efficiency/innovativeness) or the wider economic productivity the function or service may assist in the community?

Additional problems with indicators are that we often find it difficult to produce or collect any reliable or relevant data—we frequently do not collect data on the actual things governments are interested in or commit resources towards. For instance, Australian governments commit funds to improve the 'school readiness of children' in early childhood yet there is no definition of what 'school readiness' means and no data collected on indicators of readiness. It is not even clear what kind of data might be collected in the first place. In some cases, governments can produce some data but it is of poor quality, unreliable, distorted, and non-comparable over time or between agencies/units. Data comparability can be impeded when actors use different definitions of concepts or designated activities (e.g., prison rehabilitation) and complicated by different collection methodologies. Data can be captured at different times or across totally different contexts (socio-economic, regional, ethnic or indigenous communities). Delays in processing or reporting data from the front line (compilations of which can sometimes take years to see the effects) can erode the data's usefulness in judging the effectiveness of policies/programmes or in policy/programme redesign.

Making judgments on poor quality data is inherently risky and problematic. The relationship between different sets of indicators pertaining to performance (say waiting times in hospitals versus health outcomes) are not necessarily straightforward or consistent. Improvements in one measure may come at the cost of another, or the measures may have almost nothing to do with each other. Some governments in introducing performance measuring regimes have called for far too many indicators of performance (the multiple indicator malaise)—such that agencies spend inordinate amounts of time chasing data on indicators and managers do not know what expectations they are meant to be managing towards. Indicator collection can become a veritable industry, and governments or legislatures tend to multiply the number of indicators required, piling demands for more information on top of other requests (including for accountability and corporate requirements) without culling indicators to a

feasible minimum.

As a general rule, the more we collect data on discrete activities and the associated outputs of an agency the more we can produce quantifiable indicators (but subject to the above qualifications); the more we seek to collect data on outcomes and the effectiveness of programmes the more problematic will be the indicators. Indeed, many governments do not collect data on outcomes they espouse as their objectives (e.g., environmental sustainability, safer communities and improved community capabilities, and improved social cohesion). Moreover, when we ask front line staff about the appropriateness of indicators they often reply that the indicators are of little use in their operations, are burdensome to report, distorting to behaviour, and often not really getting at what they are supposed to be doing (e.g., collecting quantifiable data on numbers of publications in universities when the issue is to measure the original contributions to knowledge).

There are also special problems in both federal and unitary nations with intergovernmental transfers and shared responsibilities for policy sectors. In federal nations sub-national governments will often accept funding proposals initiated by the central government without necessarily being committed to these policy objectives. Hence, considerable gaming can occur (e.g., over precise commitments, matching funding, the timing of resource commitments—front-ending versus back-ending); and some jurisdictions can regard gaming is an expected and legitimate part of intergovernmental relations because sub-national jurisdictions may have limited rights to taxation and no guaranteed shares of the taxes raised from their own residents.

In nations like Australia and Canada there is limited evaluation of performance where funding crosses jurisdictional boundaries, because neither level has the formal powers or sometimes even the inclination to monitor and hold to account. Their legislatures rarely investigate (if ever) policies/programmes/grants that are specified by the central executive, but operationally implemented by state or provincial governments; nor do their principal accountability actors (parliamentary committees, auditors-general, commissions of audit; performance or cost commissions). And if constitutional arrangements prevent substantial institutional re-design, then it is hard to make significant political headway over performance in areas of shared responsibilities. While various occasional pleas for reform can be made by actors (such as separating roles and responsibilities, rebalancing VFI, moving to guaranteed revenue shares, transferring responsibilities entirely one level of government) they have had little practical effect. Both these jurisdictions have only just enabled their respective auditors-general to investigate and audit performance over cross-jurisdictional funding arrangements.

Myths Associated with Performance Management

There are many myths associated with performance management in the public

sector. These include that trust in government will increase or improve if governments perform better. This assumed association is neither a one-to-one relationship nor necessarily a positive relationship. Improved services can inflate community expectations and lead to disappointment and greater distrust. And as governments have attempted to show performance improvements in policy areas, it has often bred increased cynicism and disbelief.

Some believe that innovation creates better performance, but again the relationship is ambiguous. Innovations aimed at increasing effectiveness may come at the cost of efficiency. Innovations can increase cost-structures or timeliness—especially bottom-up innovations which may simply transfer transaction costs from providers to receivers. Moreover, major technological innovations can be costly investments, slow to implement, and take years to evaluate the benefits. And as Dunleavy and Carrera (2013) have argued major ITC innovations can undermine productivity in the short to medium-term, produce bureaucratic cultures of resistance, and be difficult to evaluate using a cost-benefit analysis. Similarly, many managerialists argued and believed that increased flexibilities in public management would lead to increased performance, yet when it came to demonstrating improved performance were generally unable to convincingly prove results. For instance, the Australian parliament repeatedly asked the government to produce reports evaluating the public sector reforms of the 1980s and 1990s, and although four reports were produced over a decade they were largely silent on performance improvement.

Personal incentives to improve performance are equally problematic. Rewarding high performers or high-performing units (branches, schools) can lead to the opposite effects system-wide—through discouragement of low-performers, envy, perceptions of fairness and unfairness, disputes over criteria/possibilities. It is not clear that rewarding high-performing entities actually leads to improved performance in those entities, and fortune/happenstance may have been an important contributor to enhanced performance levels. Performance pay for officials (or bonuses) again does not often lead to heightened individual or agency performance, and in their application can produce resentment, an undue emphasis on individualism, inflated exaggerations of personal contributions, team discouragement, even distorting of data and indicators (ANAO 1993). There can also be criticisms of the politicized system of reviewing and awarding performance bonuses, which may reward factors not related to performance (seniority, reputation, sensitivity, gender, regional-urban etc). This raises the whole issue of whether we can reward public sector people for good performance—it is not an easy question to answer.

There is also a powerful myth that modern public policies are evidence-based and supported by factual empirical assessments of circumstances, and that this can then (and should then) be tested to evaluate the value of the policy intervention. Performance measurement then is a form of policy evaluation. However, policies are still politically driven and governments have invested considerable 'face' in their policies, meaning that

unflattering evaluations threaten to embarrass the government or selected ministers. Under such conditions policies may evade review or be only partially reviewed to avoid criticism of the policy intent or design.

What do we get from Performance Measurement? The Tricky Task of Judging Performance

Perhaps the most significant problem with performance indicators is that the connection between the reported indicator(s) and the desired results is generally opaque, and often the subject of much disagreement (Bouckaert and Halligan 2008). There is often no clear causal link between sets of indicators and the objectives of the policy/programme. We can readily capture reams of data but much of it may not tell us much about performance; and some may mislead. Moreover, the indicators may not be controlled by the agency or programme operators; and many other factors beyond the control of the agency or programme operators may contribute to the measure. Hence, as McDavid and Hawthorn (2012) have warned, reading off judgments of performance from flawed data or from information that does not tell the overseer or reviewer what they think they are getting. They recommend developing cultures of prudence and realistic appreciations of what is possible given circumstances.

Secondly, there is the question of whether the overseers and evaluators have the requisite skills and experience to review performance (McDavid and Hawthorn 2012). Do they have sufficient content knowledge and experience with similar policies/programmes to be able to credibly evaluate similar policies/programmes? Professional training in evaluation can often be perfunctory or inappropriate. Staff in specialist review bodies quickly lose their 'groundedness' and ability to empathise with implementation realities. Experienced joint-management reviews were once fashionable (consisting of elite peer reviews by independent experts) but have since tended to fall from favour (except with project management—e.g., UK Gateway Review methodologies). Generally, the review capacities of government in judging performance-related data are not great or verifiable (independently reviewed or audited).

Then, thirdly, there is the issue of whether governments actually use performance data when making policy choices, re-authorizing programmes, re-committing resources to ongoing programmes. Arguably, only limited use is actually made of performance information by governments and relevant overseers when judging policy outcomes and re-considering approvals or extensions (and unfortunately there is almost no direct involvement by the citizenry in policy/programme review—it remains almost an entirely perfunctory executive prerogative).

The Logic of Linking Performance Measurement to Budgeting And Further Resource Allocation

There has been much recent rhetoric associated with the need to link performance measurement with budgeting and resource allocation. This invokes the notion that the

budget process can be structured to enhance performance/productivity and also demonstrate better performance in reporting. Yet, the links between budget decision-making and performance remain weak (if existent at all) and may have counter-productive or perverse consequences (e.g., an increase in prison escapes contrary to performance objectives may result in political embarrassment and the allocation of additional resources to prison managers—and skeptics may consider that there may also be an incentive to be lax with security).

Many governments have moved away from line item budgeting to framing the budget around expected results (e.g., outcomes, outputs, key performance indicators). Budget documentation has been extensively revised to provide results-based frameworks against which resources have been allocated. The impression given is that the government has weighed value for money considerations against its expected results and allocated resources accordingly. Cynics might argue that these governments have allocated their resources along traditional (legacy) lines and then presented them cosmetically in a result-based packaging. Certainly more performance information has been included in recent budget documentation (whereas little was included some decades ago) but it is not clear that an assessment of performance informs the basis for resource allocation.

If we see performance monitoring as a means of giving legislatures and taxpayers accountability information on the value for money of public provision then there are still major gaps in the governmental systems. Governments have talked about performance for over 30 years now and have accepted results-based strategic plans and budgetary documentation, but are reluctant to release detailed information and assessments of performance. Some jurisdictions provide more performance-related information in intended planning documents but such intentional commitments may not then be used for reporting purposes. Often what governments report is politically motivated, the parameters, indicators, or outputs can be changed by the executive to disguise lower than anticipated performance results.

These developments have encouraged some governments recently to talk of adopting 'performance budgeting'. But the concept is problematic for many reasons. Allen Schick began a study in the US entitled 'Does Performance Budgeting Perform?' and tried to disentangle the logics of linking the two processes together. He quickly found that performance budgeting was largely a hollow commitment and that there were insurmountable problems to be found in evaluating its performance as a new system of effective resource allocation.

Although there is currently a huge effort going on to link performance issues into the budget process, I would suggest that this trajectory is misplaced. Linking the two fields of budgeting and performance is likely to see neither achieved well—they essentially run on different and discrete logics. We may undermine the integrity of budgeting by insisting performance-related evidence is incorporated and used in some

way as a basis of decision-making; and conversely, we may undermine the integrity of performance improvement if we try to run it through the budget process. In short, I would argue that contrary to notions of moving to performance budgeting—the degree to which the budget is a suitable mechanism for examining performance is highly questionable.

Some of the more concrete problems associated with performance budgeting are: how might we reward or discipline parts of the public sector in a results-based budget framework? It is always possible to incorporate performance-related information into the budget process, but there are some real problems in using this information to formulate budgetary allocations. How might jurisdictions implement performance budgeting, would it make budgeting overly complicated and political, and what other problems would they encounter? Do we set minimal quotas, performance indicators or stretch targets? How can performance be accurately measured when line agencies will be largely in control of the performance information fed back into the process?

In the private sector there is far less argument about performance measures—but they are simpler and there is far less disagreement on the important indicators. In the public sector, the concept of performance and the nature of measurements (or findings) is highly contested and the subject of intense disputes. There is no agreed basis of comparison over time or between agencies. Also governments have to perform as well as they can in qualitative functions (e.g., economic policy advice, the assessment of the needs or training requirements of the unemployed, the facilitation of a competitive environment) or in areas of society in which they have been forced to act or take on a role (e.g., law enforcement, social assistance, treatment of the sick, housing the poor). If an agency over-performs do we reward it or cut its budget? If an agency underperforms do we censure it, reduce its funds, or allocate additional funds to help lift its performance. Is the answer different if a senior or indispensable minister is involved? In any of these cases the behavioural incentives lie essentially with the agent not the principal (as does the information). There are no easy answers to this conundrum—but perhaps we should try to address performance issues in their own terms not as part of a resourcing process. Some may respond that if we do not put performance issues into the budget then any actions that follow will not have teeth, but this suggests that the budget process can be used in such a way without generating perverse consequences, and that we cannot seek to improve performance even while leaving budgetary considerations to one side.

In some jurisdictions we are seeing governments attempting to reward or sanction delivery agencies for satisfying pre-determined performance standards. Their thinking is that delivery agencies will respond to behavioural economics and achieve higher levels of performance to receive additional performance payments for their programmes. There can also be sanction penalties associated with these initiatives, but already we know that the process of actually imposing sanctions is difficult and politically problematic for various reasons (imposing penalties on other jurisdictions, imposing

penalties on political or administrative colleagues, imposing penalties on front-line operators etc). Many public sector experts believe that sanctions do not work and that penalty provisions are deleterious to good performance or improving performance.

Conclusion

Let me emphasize I am not against performance reporting and performance management. Rather, I am concerned that many of the systems we use are partial, distortionary, ill-equipped for purpose, not good at conveying important information on effectiveness and outcomes. This is not to suggest that governments should cease attempting to improve their performance measurement and reporting regimes, but that governments should look to the bigger picture of performance improvement, the lessons of policy learning, better policy design and improved implementation. Governments collectively need to explore what factors are driving performance results (positive or negative) and use these actively in pursuing a learning agenda. I would argue that a cooperative and no-fault environment is important to encourage real improvements in performance to materialize (rather than an adversarial, insular, defensive mindset interested in 'gotcha' politics and scapegoating). Hence, a different inter-agency culture may be required, along with greater initiative from amongst those organizations and agencies most directly involved in service design and delivery.

As a general point, I would suggest simply imposing more performance reporting requirements will have limited benefit to the overall public policy process, and it may harm actual output levels and the search for greater effectiveness. Improvements in measurement, transparency (show and tell or shame) and performance reporting can only go so far. They are relatively sterile measures, often punitive in rationale, retrospective, likely to discourage innovation or risk-taking, and likely to enhance the incentives for gaming within the system as a defensive instinct. It is also onerous to comply with the accountability burden for no apparent gain in performance or community benefit.

So, how might we improve public sector performance—for a start I would recommend we move from a performance management mindset to a learning and experimental mindset (see Sable 2005). The following points of advice may enable performance improvement to become an on-going learning experience that is not threatening or punitive but constructive and rewarding.

- Importantly, minimise the scope for arguments, get agreement on the usefulness of indicators;
- Ask the deliverers and those closest to the frontline how to measure real improvements, they usually know best if they are honest;
- Build in agreed implementation schedules for performance reviewing and improvement;
- Encourage learning by doing, scope for experimentation, how are agencies cultivating their capacities, anticipating their futures;

- Avoid using narrow output reporting tables, try to frame narratives, conduct mapping and client journeys of service experiences;
- Think of ways to motivate public organizations and their employees/networks—invest in celebrations, off-line rewards;
- Pick three most important high-level measures it would be desirable to achieve, not more;
- Use hard-edged complaints systems with real sanctions, and with the CEO personally accountable; and
- Engage the public in assessing services, ask for community assessments, as well as inputs from outreach agencies and constructive reviewers.

With these cultural and behavioural factors in mind, there are great possibilities for oversight bodies and legislatures to work with providers and delivery agencies to assess delivery chains to explore where significant improvements can be made and where governments could make better investments of resources. But this would require the development of better relationships between these bodies, based on mutual endeavour, cooperation and a common interest in resolving problems. Many of these oversight bodies as institutions stemmed from the hostile scrutiny accountabilities coming out of the nineteenth century, highly political, adversarial and blame-oriented in character. Initiatives such as joint performance auditing, constructive joint-management reviews especially of projects, learning from doing and from bottom-up experience, all have potential to improve policy design and delivery. Legislatures could also be encouraged to take on a more managerial oversight role rather than a tribunal or inquisitorial role. There is much scope for upper houses and parliamentary committees to move in this direction and undertake some of this work as well as for other specialist policy-review bodies, think tanks, and even community review mechanisms.

References

- ANAO (1993). Australian National Audit Office, Performance Pay—performance appraisal and pay in the Australian Public Service, Audit Report 16, December, Commonwealth Government Printer, Canberra.
- Dunleavy, P. and L. Carrera (2013). *Growing the Productivity of Government Services*, Edwards Elgar, Cheltenham UK.
- Lindquist, E., Vincent, S. and J. Wanna (eds) *Putting Citizens First: engagement in policy and service delivery*, ANZSOG, ANU e-press, Canberra,
- McDavid, J. and L. Hawthorn (2012). *Program Evaluation and Performance Management*, Sage, Thousand Oaks.
- Radin, B.A. (2006). *Challenging the Performance Movement*, George Washington Press, Washington DC.
- Sable, C. (2005). 'Principal-Agent Governance: experimentalist organizations, learning and accountability, Wetenschappelijke Raad voor het Regeringsbeleid, The Netherlands.
- Schick, A. (2005) 'Does Performance Budgeting Perform', ANZSOG Presentation, ANU, Canberra.
- Wanna, J., Butcher, J. and Freyens, B. (2010). *Policy in Action*, UNSW Press, Sydney.