OCCASIONAL PAPERS

No. 9

PERFORMANCE MANAGEMENT IN GOVERNMENT

CONTEMPORARY ILLUSTRATIONS
PERFORMANCE MANAGEMENT IN GOVERNMENT

CONTEMPORARY ILLUSTRATIONS

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Pursuant to Article I of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

- to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
- to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and
- to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original Member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became Members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971), New Zealand (29th May 1973), Mexico (18th May 1994), the Czech Republic (21st December 1995) and Hungary (7th May 1996). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention).
FOREWORD

This Occasional Paper contains papers prepared as part of the Public Management (PUMA) Committee’s work on issues of managing performance in government organisations.

These papers further discuss issues explored in PUMA Occasional Paper No. 3, *Performance Management in Government: Performance Measurement and Results-oriented Management*, concerning two key related elements in public management reforms in OECD Member countries, namely greater devolution of decision-making authority to government organisations and an increased emphasis on their performance. They also further develop issues discussed in PUMA Occasional Paper No. 5, *Performance Measurement in Government: Issues and Illustrations*.

Tim Irwin wrote his paper during an internship he spent in PUMA while a graduate student of Princeton University, USA. It analyses the theoretical underpinning of the New Zealand public management reforms, particularly the distinction made in New Zealand between accountability for outputs of public services as opposed to outcomes. While there are numerous articles and publications describing the New Zealand reforms there are few which have sought to evaluate this underpinning.

The paper by Dr. Yvonne Fortin focus on performance management in a situation of greater management devolution of autonomy and was written when the author worked in PUMA while on leave from the French *Fondation nationale des sciences politiques*. Sweden, with its long-established autonomous agencies, has a different approach to management devolution and autonomy than most other Member countries, and is a country where devolution long preceded a focus on performance.

The papers by Mr. Walter Groszyk, Dr. Sylvie Trosa and Mr. Bram Sorber were presented to a PUMA meeting on Using Performance Measures in Government held in November, 1995. They illustrate significant performance management developments in three Member countries.

The paper by Mr. Walter Groszyk discusses the US Government Performance and Results Act, which mandates the preparation of strategic and annual performance plans by federal departments and agencies, the reporting of actual performance against plans and the development of linkages with budget appropriations. As such it represents a legislative intention to fundamentally reform the way in which government programmes are managed.

The paper by Dr. Sylvie Trosa outlines Australian experiences in benchmarking of performance of public sector organisations, a performance management tool that is attracting increasing interest in many Member countries, which includes interest in international benchmarking of performance. The paper illustrates both the benefits and pitfalls of performance benchmarking.

The Netherlands’ experiences in the development of performance indicators are described in the paper by Mr. Bram Sorber. The paper illustrates the many bureaucratic and methodological problems
that must be overcome to develop useful performance measures, as well as the need for an appropriate political and management environment if these are to be overcome.

Taken together, these papers illustrate the increasing momentum of performance management developments in Member countries. They also raise many common issues which must be faced when developing a focus on performance as part of overall public management reform.

The papers were edited by David Shand, Principal Administrator, who is responsible for PUMA’s work in the area of performance management.

The papers are published on the responsibility of the Secretary-General. The views expressed are those of the authors and do not commit or necessarily reflect those of governments of OECD Member countries.
# TABLE OF CONTENTS

Foreword ................................................................................................................................................ 3

Chapter 1
An Analysis of New Zealand’s New System of Public Sector Management
by Tim Irwin ........................................................................................................................................... 7

Chapter 2
Autonomy, Responsibility and control — The Case of Central Government Agencies in Sweden
by Yvonne Fortin ................................................................................................................................... 33

Chapter 3
Benchmarking in Public Sector Management
by Sylvie Trosa and Suzanne Williams ................................................................................................ 45

Chapter 4
Implementation of the Government Performance and Results Act of 1993
by Walter Groszyk ............................................................................................................................... 71

Chapter 5
Implementation of Government Performance and Results Act 1993
Commentary by Richard Boyle ............................................................................................................. 87

Chapter 6
Developing and Using Performance Measurement: The Netherlands Experience
by A. Sorber ........................................................................................................................................... 93

Chapter 7
Developing and Using Performance Measurement: The Netherlands Experience
Commentary by Steve Nicklen ................................................................................................................ 107
Chapter 1

AN ANALYSIS OF NEW ZEALAND'S NEW SYSTEM OF PUBLIC-SECTOR MANAGEMENT

by

Tim Irwin

Introduction

Over the last decade, the New Zealand Government has put in place a new system of public-sector management. Many of the elements of the new system are similar to those that have been introduced, or are being introduced, in other OECD countries. The government has given managers more freedom to choose inputs, for example, while paying more attention to results. The New Zealand system differs, however, from those in other countries in at least three respects:

♦ First, it makes the heads of government departments formally accountable to ministers by the use of written agreements and ex-post performance assessment.

♦ Second, it emphasises a distinction between two sorts of interest the government can have in an organisation, an “ownership interest” and a “purchase interest”.

♦ Third, it draws a sharp distinction between outputs and outcomes, the providers of government services being held accountable only for the former.

This paper discusses some of the salient issues arising from these three distinctive aspects of the New Zealand system. It focuses on what is sometimes called the “core” public sector of central government; it relates only indirectly to the issues surrounding corporatisation and privatisation of government-owned businesses, and it does not consider local government.

Part II of the paper provides a schematic sketch of the New Zealand system of public-sector management. Part III then compares this system with Australia’s and the United Kingdom’s — two countries with similar systems of government which have undertaken comparable reform programmes — to identify some of the distinctive features of the New Zealand system. Having done that, the paper examines in more detail the three aspects of the New Zealand system mentioned above.

Part IV considers the use of contracting and ex-post performance assessment as mechanisms for improving government agencies’ performance. It argues that these mechanisms probably help but cannot be relied upon alone to get good performance from government agencies. They need to be buttressed, as they are in New Zealand, by less formal mechanisms.
Part V discusses the distinction between the interest the government has in an organisation as a purchaser of the organisation’s services and the interest it has in an organisation as its owner. In particular, it discusses the implications of this distinction for the government’s monitoring of departments. It argues that, although a government does need to consider both its ownership and its purchase interest in its departments, it generally should not take a traditional profit-maximising approach to ownership monitoring and should not operationally separate purchase and ownership monitoring.

Part VI discusses the distinction between outputs and outcomes—the distinction roughly between the goods and services the government buys and the goals it hopes to achieve by buying them. It considers whether the government, as a purchaser of goods and services, should hold its suppliers, including government departments, accountable for outputs or outcomes. It argues first that there is a spectrum of states of affairs, ranging from clear outputs at one end to clear outcomes at the other, rather than a dichotomy. It argues, second, that in most—though not all—cases it is better to hold organisations accountable for things toward the output end of the spectrum. At the same time, it notes that governments have a strong interest in employing systems that provide for routine and rigorous evaluation of the effects of outputs on outcomes, and argues that the New Zealand system is weak in this respect.

How The New Zealand System Works

In broad terms, a public-sector management system, in the sense of the term used here, is the system the government uses to plan its activities and to attempt to ensure that its plans are carried out. The system has both formal and informal elements, the latter including peer pressure, a culture of professionalism and loyalty, and so on. This section discusses the formal elements of New Zealand’s system of public-sector management. The description is highly simplified, ignoring for the sake of clarity some of the complications of the formal system and the ways in which the actual workings of the system depart from the workings implied by its formal structure. And it is important to keep in mind that the formal system is not the sole instrument used to achieve goals. Boston (1991b) and some of the other papers in the same volume and OECD (1993) provide institutionally richer descriptions of the system.

To understand the system, it helps first to know the fundamental accountability relationships that it is based upon. Essentially, the heads of government agencies are each accountable to one or more ministers. When a department provides services for more than one minister, the department is accountable to each minister for the services provided to him or her, but responsible for the management of the department to just one, designated, minister. Ministers, in turn, are accountable to the legislature for their performance. The legislature, Parliament, is ultimately accountable to the electorate.

The roles of ministers and government agencies in the system have been distinguished as follows (New Zealand Government, 1990, pp. 49-50). Ministers choose which outcomes (i.e., goals) to pursue and then choose which outputs (i.e., goods and services) to produce or buy to meet those goals. Government agencies—and other organisations from whom the government makes purchases—are thus accountable to ministers for the production of outputs. Ministers are accountable to Parliament for the choice of outputs and for the choice and achievement of outcomes.

The formal public-management system can be broken down into three components: (1) ex-ante agreements setting out plans for the period ahead, (2) interim monitoring and ex-post reporting on the
extent to which the plans are being, or have been, achieved, and (3) the assessment of performance and the application of rewards or penalties.

**Ex-Ante Agreements**

Each year, about three months before the budget, the Executive sets out its strategic priorities for the budget and its long-term fiscal-policy goals. Recently, the Executive has also published a set of “strategic Result Areas” in which it lists the key contributions the public sector is intended to make over the next three years to the achievement of the Government’s long-term goals (New Zealand Government, 1995). In the budget, the Executive then sets out detailed expenditure plans for the year ahead. Portfolio by portfolio, ministers specify:

- what groups of outputs they intend to buy from government departments, non-departmental government agencies, and other organisations;
- what investments they intend to make in government agencies; and
- what grants, benefits, debt-repayments, and other payments they intend to make (see New Zealand Government, 1994a, p. 5 for details).

The budget provides a description of each group of outputs ministers intend to buy, along with various measures of their quantity, quality, and cost, and a statement of the outcomes to which they are intended to contribute. The budget focuses on outputs. Although departments estimate the value of their consumption of various categories of resource, such as personnel and operating expenses, the budget does not prescribe what types of input are to be used in the production of goods purchased or provided by the government; it leaves this decision to agencies. Moreover, though the budget states the outcomes to which the outputs are intended to contribute, it does not attempt to provide a justification for the purchase of outputs, or to describe the linkage between outputs and outcomes.

Parliament must approve the amounts of money ministers want to spend on each output group, as well as the other payments they wish to make. In a sense, the budget represents an agreement between the ministers and the legislature. In recent years, the government party has held a majority of seats in Parliament, and Parliament’s approval of the budget has been a formality. With the advent of a new electoral system based on proportional representation it is possible this will change.

Paralleling this agreement between the Executive and the legislature are more detailed performance agreements between ministers and the heads of the government departments that report to him or her. The agreements take the form of contracts, signed by both the minister and the chief executive, as the agency heads are known. The agreements themselves do not have the force of legal contracts, since the minister and his or her department are considered part of the same legal entity, but they do bear upon the legal employment contracts of the heads of departments.

These performance agreements distinguish between what they describe as the Government’s purchase interest in the chief executive’s department and its ownership interest in it. (Although the government buys services from organisations it does not own, it owns all its departments. Ministers may have contracts with agencies other than government departments, but these do not fall under the formal system of performance agreements between ministers and departmental chief executives).
The main part of the performance agreement specifies some “key results” that the chief executive should give priority to achieving. The key result might be the delivery of one or more important outputs linking to the Government’s key strategic objectives, as further discussed below, the improvement of the department’s accounting systems, the achievement of some major departmental reorganisation, or some other important goal.

The performance agreement includes as an appendix a listing of ownership requirements. It also makes reference to the purchase agreement between the minister and the chief executive. The purchase agreement is a separate document that specifies the outputs the minister wishes to buy from the chief executive. It has a format consistent with the corresponding budget information and must be consistent with the budget, but is generally more detailed. The ownership requirements are intended to specify how the government’s interest as the owner of the department will be advanced.

**Ex-Post Reporting**

Under the system, chief executives provide reports to their ministers that compare actual performance with planned performance, with respect to both the ownership and the purchase interests. Progress reports are prepared through the year, with a final report at the end of the financial year. Most departments present interim financial reports every month; very small departments can present them quarterly. All departments present quarterly reports of output delivery.

The Executive, in turn, provides various reports on its performance to Parliament. It provides financial statements for the government as a whole, which give information on the government’s total income and expenses, its cashflows, and its balance sheet. The Executive also provides reports on outputs it has bought from nondepartmental sources and reports of departmental performance, which include both financial statements and “Statements of Service Performance” detailing departments’ success in providing the outputs they had agreed to produce. The Statement of Service Performance is audited by the Controller and Auditor-General along with the financial statements.

The legislature can and does question ministers on their achievement of outcomes, and the government provides statistical information on a wide range of variables relevant to the achievement of outcomes. But there is no formal system by which ministers report their success in achieving outcomes to Parliament.

**Performance Assessment and the Application of Rewards and Sanctions**

The performance of chief executives is assessed at the end of each financial year. The assessment, like the performance agreement, is confidential. Although the performance agreement is between the minister and the chief executive, the chief executive’s performance is assessed not by the minister but by the State Services Commissioner, taking into account the minister’s views (Boston, 1991b, p. 100). This arrangement, in which chief executives are accountable to one person but assessed by another, may reflect a tension between competing desires to make chief executives more accountable to ministers and to maintain the political neutrality of the public service.

The system includes two formal mechanisms for encouraging good performance. First, chief executives’ salaries may vary according to the State Services Commissioner’s assessment of their performance. Second, chief executives are employed on fixed-term contracts, which may or may not be renewed, depending on among other things the State Services Commissioner’s assessment of their
performance. Chief executives are formally employed by the State Services Commissioner, but Cabinet must approve the appointments.

Although Parliament reviews the Statements of Service Performance and questions and criticises the Executive’s performance, it does not undertake any formalised assessment of ministers’ performance, comparable to the assessment of chief executives. Nevertheless, the Executive as a whole must have the support of a majority of Parliament if it is to govern.

Some differences between the New Zealand System and those of the United Kingdom and Australia

The United Kingdom and Australia are two other OECD countries which have forms of government similar to New Zealand’s and which have introduced comparable reforms of public management (see Kemp, 1990 and Keating and Holmes, 1990). It is useful to compare and contrast some aspects of the systems, to identify the distinctive aspects of the New Zealand reforms.

In all three countries, controls on the inputs used by government agencies have been relaxed. In all three, greater attention has been placed on measuring agencies’ performance. And, in all three, accounting systems have also been changed, so as to provide better information on the costs of activities. In Australia, as with New Zealand, the scope of the reform has included the entire public sector. In Britain, the reforms have so far mostly taken place within “agencies” delivering services to the public, rather than “departments” providing services directly to government. Agencies do, however, count for a large part of the British public service (60 per cent in terms of employment, according to Schick, 1996). (Note that in the rest of this paper “agency” does not have this special meaning; it refers to any government organisation.)

In comparing New Zealand’s reforms with others, commentators sometimes emphasise the comprehensiveness of the New Zealand reforms or their intellectual underpinnings. As mentioned earlier, this paper, however, focuses on three differences in the specific design features of the New Zealand system: the use of contracting and performance assessment to formalise chief executives’ accountability to ministers, the distinction between purchase and ownership interests, and the role of outputs and outcomes.

New Zealand’s system makes the heads of government departments formally accountable to ministers by the use of performance agreements and ex-post performance assessment. Chief executives are not just set performance targets reflecting ministers’ goals; they sign written agreements with ministers specifying what they will do. Further, their success in executing their targets is assessed at the end of each financial year, and their pay may be varied accordingly. In Britain, there are agreements between ministers and the chief executives of “agencies”, but not between ministers and the permanent secretaries of government departments. In Australia, there are no formal agreements or formal assessments of performance of chief executives. However, performance information (generally a mixture of outputs and outcomes) is reported in Program Performance Statements which are part of the budget documentation.

The New Zealand system is also different in its emphasis on the distinction between the government’s purchase and ownership interests in organisations. The split is not just a matter of theory; it finds it way into budgeting. For example, departments may sell services to various ministers, each of whom has a purchase interest in the department, but only one of the ministers is responsible for the government’s ownership interest in the department. Capital contributions are
sought by the minister responsible for the ownership interest, even if he or she is not the purchaser of the outputs the capital injection relates to (see New Zealand Government, 1994a). According to Ball (1992a, p. 18),

Recognition of these two interests [purchase and ownership] was crucial in developing the structure of financial management reform, [and] it had implications for budget and appropriation processes, for the manner of departmental operations, and for appropriate forms of reporting.

By contrast, neither the Australians nor the British appear to place any great weight on the distinction, though they may mention it (in the British context, see Kemp, 1990, p. 190).

Finally, the New Zealand system is distinctive in that it draws a sharp distinction between outputs and outcomes and holds chief executives accountable for outputs but not outcomes (e.g. Scott and Gorringe, 1989; Scott; Bushnell and Sallee, 1990; and Ball, 1992a). In the United Kingdom and Australia, the distinction does not seem to be so sharply drawn, and when it is drawn it is sometimes argued that the performance of managers should be measured in terms of outcomes as well as or instead of outputs. Describing the Australian reforms, Schick (1996) notes that the program performance statements generally report on outputs rather than outcomes, and still favour process-related objectives, suggesting that output and outcome measures are both used, albeit with greater emphasis on the former. Describing the United Kingdom’s reforms, he notes that “Treasury may also become involved [in the setting of performance targets] if it feels that the targets are not sufficiently robust or do not focus on outputs and outcomes.” Later, he refers to a review which notes, apparently with dissatisfaction, that most targets still focused on outputs rather than outcomes (Schick, 1996).

Contracting and Performance Assessment

Discussing how chief executives are made accountable to ministers in the New Zealand system, Ball (1993, p. 4) writes:

Formal contracts are central to this management accountability. A common issue for accountability is how to ensure the agent, in this case the chief executive, carries out the principal’s, i.e. the minister’s, wishes. In many other jurisdictions this is achieved by the appointment by politicians of departmental heads and senior managers who hold compatible views. .... The [New Zealand] reforms preserve a non-party political civil service ....

Hence, in the New Zealand model, formal contracts with incentives are used to achieve the alignment of interests.

A critical question is whether or in what circumstances a contractual system, such as New Zealand’s or the one used in the United Kingdom in the management of nondepartmental agencies can work to improve the performance of chief executives. That is, is it better than a system such as Australia’s or the departmental part of the United Kingdom’s, in which the performance relationship between ministers and departmental heads is more informal?

I argue below that written agreements and ex-post performance assessment probably help. Nevertheless, there are limits to the effectiveness of both mechanisms, and their use should not
distract attention from other means of improving performance. The point is not that the New Zealand system now ignores these other mechanisms; it is rather that public-sector reformers anywhere need to keep in mind the limitations of written agreements and ex-post performance assessment. While the New Zealand government understands these limitations there is little discussion of them in much of the literature on the New Zealand reforms.

At a basic level, New Zealand ministers now have relatively complete and precise information about what their departments do and how much it costs. Before the development of written agreements (and output-based appropriations), even ministers with some experience may not have known the full range of their department’s activities. This improved information probably facilitates better ministerial management and budget decision-making.

Second, government agencies themselves probably have a clearer understanding of their outputs, which by itself may be a spur to greater productivity (Drucker, 1992, pp. 80-103).

Third, written agreements probably give both ministers and departments a more consonant understanding of what the department is expected to do in the year ahead. With such an understanding, ministers have a stronger basis from which to make legitimate complaints about poor performance at the end of the year and chief executives have a stronger basis from which to defend themselves against unreasonable complaints. Similarly, performance assessment may encourage both the minister and the chief executive to think reflectively and systematically about how they can together improve the value of the services supplied by the department. Boston cites evidence that chief executives find formal performance assessment useful (1991b, p. 101).

Fourth, in addition to the formal incentives created by the system, the clear specification of expected performance may create informal incentives for good performance. In New Zealand, ministers and departments can now compare different departments’ daily “charge-out rates” for policy advice (Ball, 1993) and other outputs produced in more than one department. A desire to be considered a good performer may, therefore, encourage departments to lower their prices or demonstrate that their prices are somehow justified. Moreover, having ministers sign purchase agreements with departments may cause them to rethink their role, prompting them to think of alternative sources of supply, and disinclining them to act only as their department’s representative in Cabinet.

Finally, once a department has committed itself to delivering various outputs, the minister and other interested parties who are unhappy with the department’s performance can criticise the department for failing to meet the commitment — even if their dissatisfaction arises from other aspects of the department’s performance. Failing to perform according to the performance or purchase agreement thus makes the department vulnerable to criticism, and the department has an incentive to carry out its agreed tasks.

**General Limitations on the Efficacy of Contracting and Performance Assessment**

Despite these likely advantages, contracting and performance assessment have limitations that need to be acknowledged.

The limitations of contracting arise because contracts are, in general, necessarily incomplete (Williamson, 1989; Klein, 1985). It is usually impossible to specify exhaustively the features of the product the buyer wants from the seller. And, in contracts in force over an extended period, it is
generally impossible to specify what each party should do in each possible circumstance that may arise: the contingencies are just too numerous. Even when it is possible to specify in advance what is to be done, it may be advantageous for the parties to defer these decisions until after they have entered into a contract (Simon, 1951). Written contracts may even be relatively unimportant. Llewellyn, a legal theorist, has described a contract as “a framework which almost never accurately indicates real working relations, but which affords a rough indication around which such relations vary, an occasional guide in cases of doubt, and a norm of ultimate appeal when relations cease in fact to work” (1931, p. 737).

Nevertheless, buyers often do get what they want. Contracts work, incomplete though they often are. How do they work despite their incompleteness? They work because the contracts are only one of the mechanisms that are used to help buyers get what they want from sellers — and in some cases, such as employment contracts, to help sellers get what they want from buyers (see e.g. Barzel, 1982; Klein, 1985). Buyers, to consider just one side of the transaction, also consider

♦ the reputation of the seller;
♦ the seller’s professional qualifications and certification;
♦ the seller’s interest in winning ongoing business and the availability of alternative suppliers; and
♦ the seller’s interest in doing the job well just for the sake of doing the job well.4

Because of this, sellers often give buyers the products they want even though the written contract does not require it.

The first lesson that emerges from this is that governments must use other mechanisms in addition to written agreements to ensure that their agencies work well. They should try to make use of and foster norms of professionalism, loyalty, and public service. They should appoint highly skilled people to run their agencies. They should make the supply of services contestable where that is possible and be prepared to switch suppliers when the incumbent supplier performs poorly — something that formal contracting greatly facilitates but does not by itself bring about.

This suggests that contracting and performance assessment may be helpful but insufficient — and this is the tentative conclusion of the paper. Evidence that they may actually be counterproductive in some cases needs, however, to be mentioned. The use of monitoring — such as detailed written agreements and performance assessment — is in part a sign that agents cannot be trusted to do the job well on their own and need to be monitored and then rewarded or sanctioned accordingly. In some cases, this signalling of a lack of trust can reduce agents’ internal motivation to work well (see Frey, 1993 and also Williamson, 1993). Similarly, the payment of bonuses for good performance may lead agents to feel that their job is not intrinsically rewarding, again reducing effort (Kohn; 1993).

Problems Specific to Government

In addition to the general limitations on contracting and assessment as means of improving performance, there are some problems specific to governments or the New Zealand system.
Notably, ministers may have little interest in assessing the performance of their agencies, unless the agencies’ activities are politically salient or are provided directly to them, as in the case of policy advice. Parliament may also be less interested in pursuing performance issues rather than other issues in its review of department or agency operations, thus reducing pressure on ministers to assess performance. There is a risk therefore that their monitoring turns out to be perfunctory and the effects of monitoring on performance weak. Departments might be able to set themselves easy targets, and in any case expect few problems should they fail to meet even these targets. Although monitoring by central agencies can help, it cannot solve the problem: the central agencies know much less than the departments about the latter’s outputs, and their monitoring can be influential only to the extent that ministers have an interest in what is monitored. Moreover, even if politicians have an interest in monitoring, they may lack the experience and the resources to scrutinise their departments’ performance. One lesson of the New Zealand experience is that ministers may need to use new sources of advice to monitor their contracts with their agencies (Steering Group, pp. 64-67).

Assuming ministers did monitor departments’ performance, the question still remains of the adequacy of the incentives the New Zealand system generates for good departmental performance. What happens to a department, for example, if it doesn’t deliver its promised outputs? In its formal mechanisms, the system places considerable weight on the chief executive’s performance assessment. Good departmental performance is rewarded by giving the chief executive a good performance assessment and perhaps a bonus. For a chief executive supplying a large range and volume of outputs, the overall assessment of performance therefore requires, in theory, a complicated weighting of the importance of different outputs. In New Zealand the specification of chief executives’ “Key Result Areas” gives some help in weighting the importance of different outputs by listing high priority areas of activity. In practice, the State Services Commission distinguishes in its assessment of chief executives between those things chief executives are to be held “personally” accountable for — such as key outputs and organisational goals — and other, more routine, aspects of his or her department’s performance (State Services Commission, 1994). Therefore, it seems unlikely that performance assessment places much pressure on chief executives to produce routine outputs efficiently.

How serious this problem is remains unclear. Noncontractual mechanisms for getting good performance — such as the appointment of good chief executives, peer pressure, contestability — may be enough to overcome this problem. But the government could also make greater use of systems that do not require making judgements about the overall performance of a department’s or chief executive’s performance. At present, for example, whether departments receive appropriations for the supply of outputs in a given year is generally independent of whether they supply the outputs in that year; departments’ performance might be improved by making payment contingent on the delivery of outputs, a system that has been discussed in New Zealand but not generally implemented (Ball, 1992a).

In summary, then, public management is probably best served, as the New Zealand Government recognises, by a combination of formal contracting and performance assessment on the one hand and strong informal mechanisms for improving performance, on the other.

Purchase and Ownership

A government may purchase goods and services from an organisation it does not own. The New Zealand Government, for example, buys services for “families in need of support” from churches, voluntary organisations and other privately owned agencies (New Zealand Government, 1994,
In such a case the government can be said to have a purchase interest in the organisation. As a purchaser the government’s interest is much the same as any customer’s: it cares about the quantity and quality of the goods, when and where they are delivered, and their price.

The Government may also own an organisation without making significant purchases from it. The New Zealand Government, for example, owns a coal-mining company. In such a case, the government can be said to have an ownership interest in the organisation. As owner its interest may be much the same as any other owner’s. Essentially it has an interest in maximising the discounted value of the agency’s future stream of profits, adjusted for the expected variance of that stream of profits. As a result it wants to know what profits it is making on the capital it has invested in the agency and how those profits compare with what it could make if it were to invest the capital somewhere else. It wants to know about factors that impinge on the agency’s expected future profits — such as the quality of its management, management practices and staff, as well as its physical assets. And it wants to know about factors that affect the expected variability of the agency’s profits such as the agency’s area of business.

The appropriate type of ex-ante budgeting and ex-post financial reporting depends on the nature of the government’s interest in the agency. As a non-purchasing owner the government has an interest in seeing financial statements like those presented to a privately owned firm’s shareholders. As a non-owning purchaser the government has an interest in statements specifying the nature of the outputs it is offered (ex ante) and has received (ex post).

The difficult questions arise when the government has both a major purchase and a major ownership interest in an agency, as is typically the case with government departments. In New Zealand most departments sell most of their output to Government and all are government-owned. Here, the New Zealand government, unlike others, continues to employ the purchase-ownership distinction.

Writing about this case, McCulloch and Ball (1992) state that:

These different interests of the Crown in departments (as owner and as purchaser) imply different approaches to performance measurement. The ownership interest is likely to require the same types of financial reporting as are generally accepted practice in the private sector....

The purchase interest is likely to require the same types of information as private sector sales contracts: quantity, quality, time and place of delivery, and price.

For the 1994/95 year, the New Zealand budget is split in two main parts that reflect the distinction. The Estimates of Appropriations detail the Government’s purchases from departments and other organisations, while Departmental budgets present financial statements for each department — operating statements, cash-flow statements, and balance sheets (New Zealand Government, 1994a and 1994b, respectively). The Estimates of Appropriations give performance measures normally relevant to purchasers — the quantity, quality, and cost of outputs — while Departmental budgets give performance measures normally relevant to owners — sales other than to the government, the “liquid ratio”, “average debtors outstanding”, “surplus/(deficit) operating activities”, and so on.

The key question is whether it is helpful for the government to distinguish purchase and ownership interests in its monitoring of a department that sells all or most of its output to the government. It is conceptually possible to distinguish the government’s ownership and purchase
interests in such a department. However, the argument of this section is that the government should not pursue its ownership and purchase interests separately and should not operationally separate the ownership and purchase monitoring of such a department because of the risks of decisions which will not improve the organisation’s efficiency. Rather, it should pursue a single goal that might be described as getting the best value for money from the department over the long term.

To consider this issue, it helps to look at how ownership and purchase monitoring work in other contexts. In most nongovernmental markets, organisations are monitored both by their purchasers and by their owners. Purchasers pressure the firm to lower prices or improve quality and compare the organisation’s products with those available from any competitors. Owners pressure the firm to increase profits, by cutting costs and increasing sales.

When consumers can observe the value of a firm’s product and can choose between different suppliers it is usually not in the interest of owners simply to raise prices and lower quality. The owner does better to produce more efficiently. Similarly, with strong ownership monitoring firms can respond to consumer pressure for lower prices and higher quality only by becoming more efficient, not by generating lower profits. Thus in a competitive market the joint application of ownership and purchase monitoring tends to encourage organisations to produce more efficiently.

The application of this sort of purchase and ownership monitoring runs into problems in situations where consumers cannot easily discern the value of the output they consume or cannot easily switch suppliers. In these cases, the easiest way for the supplier to increase its profits is to raise the per-unit price of its output or lower the quality of that output. In such a situation, traditional ownership pressure is unlikely to improve the organisation’s efficiency. Rather, increases in profits are likely to come at the expense of reductions in the value of the output to purchasers. Scott, Bushnell, and Sallee make a similar point (1990, p. 157). They note that “When it is able to set its own prices uncontested, a department could achieve any target rate of return regardless of its efficiency.”

Unfortunately, this is the situation that characterises many of the transactions between governments and their departments. As purchaser, the government is often unable to measure accurately the value of the services it is buying; that is, it cannot precisely measure the degree to which its purchase interest is being satisfied. Putting pressure on the department to increase profits is more likely therefore to lead to reductions in the value of its services than to improved efficiency. As a result, the government should not pursue its ownership interest in a department in the way that a typical firm owner would pursue its interests in the firm.

Operationally, moreover, the government should not separate ownership monitoring from purchase monitoring. As owner, the government needs to pressure the department, not simply to maximise profits, but to achieve the particular level of profit that is consistent with the cost of the capital the government has invested in it. But it needs to ensure that the department does not increase its profits to that level simply by taking actions that reduce the value of the department’s output to the government as purchaser. A unified monitoring system probably has a better chance of ensuring that.

The single-minded pursuit of purchase interest is probably less risky than the single-minded pursuit of ownership interest. The government can more easily measure the degree to which its ownership interest in a department is being satisfied: although there are problems in the valuation of capital, it can get at least a rough measure of the surplus of revenues over costs. Thus there are fewer opportunities for departments to satisfy purchase interests by sacrificing ownership interest. Nevertheless, the vigorous pursuit of purchase interest would entail some risks if it were done
separately from the pursuit of ownership interests. For example, the department might be able to lower the costs of the services it is currently providing by incurring long-term or contingent liabilities; this would be good for the government as purchaser, but bad for the government as owner. These risks would, again, be minimised by operationally combining the purchase and ownership monitoring; in that case, the single monitor of performance could, for example, seek improvements in the satisfaction of the government’s purchase interest while also checking that they weren’t achieved at the cost of the government’s ownership interest.

I conclude, therefore, that the government should not pursue its purchase and ownership interests in a department separately. Purchase and ownership monitoring should be done jointly and they might more helpfully be described as long-term value-for-money monitoring. Before leaving the subject, it is worth making two subsidiary remarks.

First, the difference between the value of traditional ownership and purchase monitoring in government and in many private-sector situations does not arise simply because the government is both owner and purchaser. The argument outlined above implies that if the government could observe the value of output accurately a combination of traditional ownership and purchase monitoring might work. Note also that the difference in the value of profit-oriented ownership monitoring does not arise out of any fundamental difference between the government and the private sector. The argument is not that the public sector is special in some sense. Traditional profit-oriented ownership monitoring is likely to be inappropriate for company divisions that provide in-house services with difficult-to-measure-benefits, such as strategic business advice. From a social perspective it may also lead to problems when applied to privately owned firms with monopoly power.

Second, although some descriptions of the New Zealand system emphasise the similarities, not the differences between traditional ownership monitoring and departmental ownership monitoring, the New Zealand Government does not in practice monitor and pressure government departments in a traditional profit-seeking way. It collects information on the surpluses (or profits) they make but does not tell — or allow departments — to maximise the surplus (New Zealand Government, 1994b). At present, it tends also to emphasise the purchase interest, chief executives having specific and detailed purchase agreements but a relatively short, standard-format section on ownership requirements in their performance agreements. That section also reveals some unclarity as to what ownership monitoring should entail, including as it does some points that appear to relate to the Government’s purchase interest in departments. For example, it says that “Chief executives are expected to: ... provide the Minister with free and frank advice that is relevant, accurate, and timely” (State Services Commission, 1995). Perhaps because of uncertainty about what the “ownership interest” involves it appears to be used as a heading for issues that do not clearly fit elsewhere. For example the ownership section in performance agreements stresses the government’s collective interests in departments, even though those collective interests may be as much to do with purchase as ownership issues. Strategic management issues have also been linked to the government’s ownership interest, even though the strategy may have as much to do with purchase as with ownership issues (State Services Commission, 1994). However officials appear to recognise the desirability of clarifying the role of ownership monitoring and have been working toward that end (State Services Commission, 1994).

Outputs and Outcomes

New Zealand’s system, as we have noted, emphasises the measurement of agencies’ outputs. In particular, chief executives are intended to be held accountable for the production of outputs and not
the achievement of outcomes. This section considers whether this is the best way for the government to manage its departments or whether more emphasis should be placed on outcome-based contracting. The plan of the section is as follows. First, we clarify the output-outcome issue, and explain why it is a fundamental problem. Then we look at the practice of contracting by individuals for the purchase of private services to see what help that offers us. Next, we consider two ways in which the government’s problem is different from those of typical private-sector principals. With those differences in mind we suggest some principles for choosing between output and outcome contracting in government and offer two specific suggestions for the practice of contracting in New Zealand.

Clarifying the Issue

We need first to be clear about what we mean by “output” and “outcome”. Ball (1992a, p. 17) provides the following definition: “outputs are goods and services produced by departments; outcomes are impacts on or consequences for the community of the interventions of government”. Typically, then, outputs are a means to outcomes. For example, hospitals might provide the output of coronary by-pass operations in order to achieve the outcome of fewer heart attacks. An employment agency might provide the output of interviewing unemployed people and giving them information about job vacancies in order to achieve the outcome of lower unemployment.

Although the terms “output” and “outcome” are often used as though the relationship between them was one of dichotomy, it is more accurate to think of a range of states of affairs, with outputs located towards one end of the range and outcomes typically towards the other. We can arrange the relevant states of affairs along two dimensions: (1) controllability by the government department, and (2) intrinsic value to the government. Outputs have high controllability and, typically, low intrinsic value. Outcomes, on the other hand, have high value and, typically, low controllability. The following diagram illustrates the point.

In theory, the government and the department can consider striking an agreement anywhere in the controllability-value “space”. In practice, however, there will usually be few states of affairs in the high-value, high-controllability region (the upper-right region of the diagram). They also have no reason to contract over states of affairs in the low-value, low-controllability region (the lower-left region of the diagram). In general, the agreements they will choose among will lie between the top-left and bottom-right regions. Typical output agreements lie in that bottom-right region; typical outcome agreements lie in the top-left region. But output and outcome agreements cannot be neatly separated; there is a range of possible agreements lying between the two extremes.

An example may make this clearer. An employment agency has considerable control over whether it interviews the unemployed, so this state of affairs is near the output end of the spectrum (the bottom right of the diagram). But it does not have complete control even over this: it cannot ensure that the unemployed will attend the interviews it schedules, for example. Even closer to the output end of the spectrum, then, are states such as the issuing of invitations to attend interviews. Moving in the other direction, towards the outcome end (or top left) of the spectrum, we can imagine the following states of affairs: more unemployed people applying for jobs, more unemployed finding jobs, fewer people being unemployed, fewer people being poor, and so on.
Choosing Between Output and Outcome Contracts: the Fundamental Problem

Although the “output-outcome” debate is often framed in terms of the question, “should governments’ formal systems of public management focus on outputs or outcomes?” it is more accurate and probably more helpful to consider instead what part of the output-outcome range the government should focus on in managing its agencies. Should it focus on the number of invitations to interviews sent out by the employment agency? Should it focus on the number of interviews conducted? On the number of people who are referred to jobs? Or the number of people who actually get jobs? Or on the unemployment rate? Despite there being no dichotomy between outputs and outcomes it is still useful to use “output” and “outcome” to refer to states towards the ends of the spectrum and we will continue to use these terms.

The key question, then, is what type of state of affairs the government should choose to contract its agencies to provide — formally or informally — and therefore measure and hold its agencies accountable for providing.

There are advantages and disadvantages with both output-like and outcome-like contracting and measurement.

Because agencies have a high degree of control over outputs, they cannot easily make excuses for failing to provide the contracted outputs. As a result, the government will have fewer problems holding an agency accountable for delivering the outputs (Ball, 1992a and 1992b). Because an agency usually has only partial control over the achievement of an outcome, it is harder for the government to hold the agency accountable for the achievement of the outcome. The number of unemployed people who get jobs depends, for example, not only on the skill with which the employment agency carries
out its actions but on the willingness of the unemployed to seek work and the willingness of employers to take them on. If the employment agency fails to meet a job-placement target, it may easily be able to blame other factors. It can say that employers became pessimistic about the future demand for their goods, for example. (Likewise, if the outcome is achieved, the government may be reluctant to reward the agency because of doubts about the agency’s importance in achieving the god.) If the agency can always credibly avoid shouldering the blame for outcome failures, it has no accountability for them.7 Anthony and Young in a seminal textbook on management control in non-profit organisations (1988, p. 608) aptly sum up the problems with many outcome performance measures (in their terminology, “social indicator” means “outcome”):

A social indicator is a broad measure of output which significantly reflects the work of the organisation. Unfortunately, few social indicators can be related to the work of a single organisation because in almost all cases they are affected by exogenous forces; that is, forces other than those of the organisation being measured. The crime rate in a city may reflect the activities of the police department and the court system, but it is also affected by unemployment, housing conditions, and other factors unrelated to the effectiveness of these organisations...

Social indicators are so nebulous, so difficult to obtain on a current basis, so little affected by current program effort, and so much affected by external influences that they are of limited usefulness in day-to-day management...

Perhaps because of these problems, no country currently has a fully fledged outcome-based contracting and budgeting system.

However output-based contracts agreements may fail to encourage agencies to do what the government wants them to do. For in many cases the agency may be able to fulfil its output contract perfectly without doing much of value for the government. If the employment agency gives bad advice, it could interview as many unemployed people as the government had asked it to, without contributing to the reduction of unemployment. The government can write quality specifications into its output contracts, but for most services it will be impossible to write and enforce a contract that requires the agency to act exactly as the government would want it to. The problem was discussed earlier: there are too many contingencies that need to be specified, and too much that the government doesn’t know about the service. Imagine the complexity of a contract that specified the actions that the employment agency should take in all its interviews with the unemployed, taking into account the different circumstances of each interview. Or imagine the nature of the required contract with a police force.

As a result, agencies could often — if they chose — fulfil output contracts without helping the government as much as the government would want. Similarly, if the government tries to write a complex performance contract specifying various quality standards the agency may be encouraged to meet those standards without regard for the government’s aims (see, for example, Osbome and Gaebler, 1992, pp. 138-139). In the education sector the problem goes by the name of “teachers teaching the test”. An agency may, it is true, have an incentive to produce the outcomes the government desires, even if the agency’s contract does not require it to. For example, the agency may face competition for the right to supply the services to the government. But in the absence of such external pressures to perform well the output contract may not get the government what it wants.
Outcome-based contracts are less likely to cause any problem of non-relevant outputs. If they actually motivate the agency to change its behaviour, they are more likely to motivate it to behave in the way the government wants it to behave. Suppose the government wants its employment agency to reduce unemployment and nothing else. If the government contracted the employment agency to do this — and the employment agency considered the contract binding — the government could expect the agency to carry out interviews with the unemployed in a way that really did help reduce unemployment. Outcome-based contracting thus has the potential to circumvent some of the problems with output-based contracting. If it works, it gives the agency the government’s objectives — thus eliminating the need for detailed monitoring of the agency’s actions or outputs.

The “if” is crucial, though. If the outcome is not sufficiently under the agency’s control, an outcome-based contract won’t be credible, the agency will not be truly accountable for the outcome and the outcome target may not motivate the agency.

In sum then, governments face the following dilemma: output contracts may fail to get agencies to do what the government really wants it to do; and outcome contracts may fail to make agencies truly accountable for anything.

Looking at Contracting Outside the Government-Agency Context

The problem of how buyers get what they want from sellers is common to many transactions — buying a car, getting a haircut, hiring a worker, hiring a lawyer, buying inputs at a manufacturing plant; it isn’t specific to the relationship between government and its agencies. It’s useful, therefore, to consider whether output or outcome contracts are used outside government. In doing so, we are not assuming that people outside government have found perfect solutions to the problems of doing business, only that the practices they employ are usually subject to some competitive pressures — the economic equivalent of natural selection — and those that survive may represent reasonably efficient solutions to the problems that are encountered (Pratt and Zeckhauser, 1985).

The easiest way to tell what the seller is being held accountable for in the contract is to ask what the seller is paid for — is it the production of an output or an outcome? Employing this approach, we find some cases in which buyers hold sellers accountable for outputs and others in which they hold them accountable for outcomes.

Most contracts — implicit or written — appear to focus on output-like states. We usually pay lawyers for giving us advice, not for winning the case. We pay doctors for advice, prescriptions, and treatment, not for our recovery. We usually pay workers according to their inputs — e.g., the time they spend on the job — or their outputs — the number of shirts they sew — and not according to the firm’s profits. We pay for the haircut, not for whatever it is we want the haircut to help us get.

In many cases in which buyers and sellers contract on an output basis, the sellers may also provide outcome-related information to buyers to persuade them to buy. Drug companies report the results of drug tests to doctors and sometimes patients, for example, even though they sell an output not an outcome (the drug not good health). Although outcome-related information is used, the sellers are held formally accountable for outputs. In cases where the parties transact several times, buyers form judgements about the effects of outputs on the outcomes they seek and decide whether to renew the contract on the basis of those judgements. In these cases, sellers get business by contributing to outcomes, or appearing to do so, but the formal contracts may still relate to outputs.
Not all contracts outside the government are settled at the output-end of the spectrum, however. Lawyers in the United States are sometimes paid on a contingency-fee basis, earning nothing if they lose their case. Firms often pay an employment agency only if they hire a worker referred to them by the agency. House sellers often pay their real-estate agent only if the real-estate agent sells their house. Apartment renters often pay the accommodation agent only if they choose an apartment advertised by the agent. Investors may pay fund managers a percentage of the return on the funds they control. Merchant banks handling the sale of a business are sometimes paid only if they successfully broker a deal. In the labour market, some employees are paid bonuses that depend on their firm’s profits, or the profits of their part of the firm. And salespeople are often paid on commission.

In all these cases, the sellers are contracted to provide, and are held accountable for providing, something which is only partly under their control. Effective accountability requires significant control, that is, but not total control. In some cases, the motivational advantages of outcome-like contracts — arising from bringing the supplier’s goals closer to those of the buyer — may outweigh their problems.

It is not obvious why outcome contracts are used when they are, but we can speculate about the conditions in which they are more likely to be chosen (see Barzel, 1982 and Klein, 1985):

♦ First, outcome contracts are only possible when the desired outcome can be clearly specified and its attainment easily measured (such as winning a court case, selling a house, finding an apartment, hiring an employee, making a profit).

♦ Second, outcome contracts are more likely the greater the control the seller has over the achievement of the outcome (a firm’s chief executive may be paid according to the firm’s total profits, but the shop-floor worker is unlikely to be).

♦ Third, and related to the preceding point, outcome-like contracts are more likely when the seller has more control than the buyer over the risks that contribute to the achievement of the outcome, or can better respond to the risks than the buyer.

♦ Fourth, outcome contracts are more likely to be used when sellers provide a large number of services, each of relatively low value (e.g. employment agents, real-estate agents, rental agents). The reason is that outcome contracts require sellers to bear more risk than output contracts do, and those risks can be borne more easily by the seller of a large number of services.

♦ Fifth, outcome contracts are more likely when the seller knows more than the buyer about the contribution of the service to the achievement of the desired outcome (lawyers know more than clients about the effects of their services on the client’s chance of winning a case).

Finally, outcome contracts are more likely when other mechanisms such as reputation, professionalism, and an interest in ongoing business cannot be relied upon.
**Differences Between the Government and Other Buyers**

Although there are similarities between the government-agency relationship and other buyer-seller relationships, there are differences too.

First, the government may know less than other buyers about the link between the goods and services it buys and the outcomes it wants. In many cases, it is buying services — such as health services and education — on behalf of others. Since the government doesn’t consume the services directly, it cannot easily assess whether it is buying the right services. One response to this problem is for the government effectively to devolve monitoring to the consumers of the service, for example by letting them choose between different government-subsidised providers. But if the government doesn’t do this it needs, more than most other buyers, to seek out information on the links between the outputs its agencies produce and the outcomes it wants.

Second, the choice between output and outcome contracting may have implications for the respective roles of politicians and bureaucrats. These implications need to be considered in making the choice.

In an output-based system such as New Zealand’s, politicians choose which outputs the government will buy and from whom. Bureaucrats then choose which inputs to use in order to produce those outputs. The reasoning is that, given output targets, the agencies are in the best position to decide which inputs to use and controlling inputs would limit the value of holding agencies accountable for outputs. In the case of outcome-based contracting, the same logic would suggest that politicians leave to bureaucrats the choice of which outputs to produce.

An agency that was contracted to reduce road deaths should, on this logic, be given the authority to choose how much to spend on safe-driving advertising campaigns, how much to spend on road improvements, how much to spend on ambulances, and so on. An agency contracted to reduce unemployment should be given the authority to choose how much to spend on interviewing the unemployed, how much to spend training them, how much to spend on subsidies to firms to employ the unemployed, and so on.

It would not necessarily be undesirable to give agencies more control over the choice of outputs, if they were effectively held accountable for outcome targets. Perhaps a system like New Zealand’s requires politicians to make too-finely detailed decisions about the means that are used to achieve the ends they seek. Perhaps the choices about which outputs to produce should more often be made by departmental managers, who know more about the effects of outputs than do ministers, even with the benefit of good policy advice. This would, however, require politicians to delegate much greater authority over what are considered policy matters, which they may be unwilling to do. It may, moreover, be impossible for them to delegate control over all the instruments governments use to achieve outcomes, since these instruments include regulations and taxes. Reaping all the potential benefits from outcome contracting may therefore be problematic.

**Implications for Governments’ Approaches to Outputs and Outcomes**

Some tentative lessons for government-agency contracting emerge from this discussion:
Output contracts may be appropriate most of the time, but governments also need to use other mechanisms for ensuring that its agencies provide the services it wants them to, including the collection of information on output-outcome links.

When governments contract for outputs on behalf of other people, they may need to do more than other buyers to assess the links between outputs and outcomes.

Outcome contracts are possible in those cases where the desired outcome is easily defined and measured and the agency has some significant degree of control over the outcome and is able to bear the risks of not achieving the outcome.

If governments contract on an outcome basis, there are reasons for giving agencies the authority to decide on the mix of outputs contributing to the desired outcome, which would, however, increase the power of agencies over policy.

In many respects the New Zealand system stands up well in the light of this analysis. Its output-based contracts facilitate clear (if arguably, narrow) accountability and we have seen that output-based contracting for the purchase of private services by individuals does not preclude the existence of incentives to produce outcomes. Among other things the Government, as part of the planning and budgeting process, specifies its “strategic result areas” and departments have an interest in offering outputs that clearly contribute to these areas. Clear output contracts also facilitate contestability of service provision, which further encourages departments to provide outputs that help the government achieve its stated goals. But the analysis also suggests two areas in which the New Zealand Government might look for improvements in its system, beyond the planned development and improvement of its existing output contracting.

Experimenting with Outcome Contracts

First, the Government might strike some contracts further towards the outcome end of the output-outcome continuum. It is beyond the scope of this paper to say which services could be contracted for on an outcome-like basis, but the conjectured principles raised earlier are relevant. Outcome contracts are more likely to be appropriate, that is, when

- the outcomes that the government wants are easily defined and measured;
- the agency has some significant degree of control over the outcomes;
- the agency knows more than the government about the effects of outputs on outcomes; and
- the agency is able to bear the risks involved.

Since the supplier must have an important degree of control over the outcome, useful outcome contracts would probably not specify high-level outcomes such as the unemployment rate or the crime rate. Moreover, to make sure agencies were truly accountable for the performance of the contract — and not simply said to be — it would probably be desirable to link payments to the department to the achievement of outcomes.
Interestingly, the New Zealand Government does in fact employ some outcome-like contracting and some outcome-based performance measurement. Like other governments and other owners, it monitors the performance of the commercial businesses it owns according to the profits they make — something which they have considerable but not total control over. It also has a contract with the head of the central bank that requires him to keep the annual inflation rate between zero and two percent. The contract appears to work, both because the central bank seems to have considerable control over the inflation rate and because the system includes means of controlling for the effects of significant inflationary shocks outside the bank’s control, such a large oil-price rise.

The Government also uses some outcome-like performance measures in the core public sector. For example, the performance measures for a job-placement output bought from the Department of Labour include the number of “successful referrals of priority jobseekers ... into subsidised and unsubsidised employment” and specifies a target of 50000 (New Zealand Government, 1994a, p. 353). Whether this is an appropriate outcome-like target is unclear, but it does raise the possibility of striking an outcome-based contract in which the Department of Labour is paid per job placement.

The best approach may be to experiment with some outcome-based contracting in those cases where the nature of the service suggests it is possible.

*Getting Better Information on the Effects of Outputs on Outcomes*

Second, the New Zealand system of output contracting would benefit from the systematic collection of better information on the effects of outputs on the achievement of outcomes.

There is nothing original about such a suggestion, but New Zealand advisers often note the need for such information. Bushnell (1989), for example, writes “Definition of outputs, however, is merely the first step of a two-step process. The second step involves assessing the contribution of the outputs to outcomes ...” Scott, Bushnell and Sallee (1990) and Ball (1992a), among others, make the point that much policy advice is advice about the connection between outputs and outcomes. And the New Zealand Government certainly spends a significant amount on policy advice (Ball 1993, p 5).

Nevertheless, New Zealand’s public-sector management reforms have not placed much emphasis on improving the information the government receives on output-outcome links — especially from systematic, routine evaluations, as opposed to ad-hoc reviews. Speaking of evaluation, the chief executive of one department says bluntly, “In New Zealand we have not been good at it” (Laking 1993, p 6). Australia provides an interesting contrast in this respect. There, the systematic evaluation of the effect of government programmes on outcomes has been an important part of the reforms of public-sector management over the last decade. Describing the Australian reforms, Keating and Holmes (1990, p 174) say that evaluation “is a crucial element of the system of managing for results”. Mackay (1993, p 10) describes Australia’s evaluation policy thus:

*Each portfolio Minister is required to prepare a PEP [portfolio evaluation plan] and submit it to the Minister for Finance. Each PEP has a three-year coverage, relating to major evaluations — i.e. those with major resource and/or policy implications, and is intended to focus mainly on outcomes and effectiveness issues. Other evaluation activity is encouraged, and each program is required to be evaluated every three to five years.*
Although the value of expenditure on outputs and other programmes is frequently reviewed in New Zealand, there is nothing comparable to this systematic production of evaluations.

The difficulty of holding agencies accountable for outcomes needs to be acknowledged. This reflects the reasons why evaluation is difficult — namely the complexity of the relationship between outputs and outcomes. Because outcomes are affected not only by the output or programme being reviewed, good evaluations must somehow isolate the effect of the output or programmes. They may require the collection of new data, the use of control and experimental groups, or the application of sophisticated statistical techniques. They may be costly and time consuming. But given the amount of money the government spends on some programmes it seems likely that more systematic and rigorous evaluation of outputs and programme outcomes would be beneficial.

The question of information on outputs-outcomes links also relates to Parliament’s scrutiny of the Executive.

Although the Executive now provides much more detailed and more useful supporting information to Parliament on the outputs it is purchasing and on the financial position of the government (Steering Group, p 39), it does not report on its success in achieving outcomes. Instead, the Executive reports on its purchase of outputs from departments and other organisations. In reviewing ministers’ performance, that is, Parliament looks at the same sort of information that ministers consider when reviewing departments’ performance. If ministers are intended to be responsible for outcomes, there seems to be a gap in the performance management system. (This is not an original criticism: see Laking, 1993; Pallot, 1991, pp. 216-220; Scott, Bushnell and Sallee, 1990; pp. 157-158).

There are good reasons why ministers might be wary of reporting on and being held accountable for outcomes. Given our current knowledge of the world, many of the government’s possible goals — higher incomes, lower unemployment, less crime — may be only partially under its control. A recession in the rest of the world would tend, for example, to lower New Zealanders’ incomes, whatever the government did. For ministerial accountability for outcomes to work, therefore, ministers could be held accountable for outcomes only in those cases where a desirable outcome that was within their control could be identified (the number of deaths in traffic accidents might be an example).

Even if ministers weren’t held accountable by Parliament for the achievement of outcomes per se, they could provide fuller outcome-based justifications of their proposed expenditure on some or all outputs. At present, the Estimates state the outcomes to which outputs are intended to contribute, but they offer little in the way of evidence that the outputs will contribute to those outcomes (see New Zealand Government, 1994a).

Ministers could also present selective ex-post evaluations of the actual contributions of outputs to outcomes. Laking (1993, p 6) notes that ministers could be required “to prepare ‘outcome reports’ for Parliament on a cyclical basis for each of their major outputs”. There is a danger, of course, in calling for yet more information to be provided to Parliamentarians with limited time. Yet there might be benefits to sacrificing some of the detail they — not ministers — receive about the outputs ministers buy, in favour of information on the link between those outputs and the government’s goals.
Conclusions

New Zealand’s new system of public management differs in at least three important respects from those of other countries, such as Australia and the United Kingdom, who have also introduced significant reforms. Specifically, it employs formal contracting and ex-post performance assessment as means of improving the performance of government departments; it distinguishes between the government’s ownership and purchase interests in agencies the government both owns and buys services from; and, in measuring the performance of government agencies, it focuses on outputs not outcomes. This paper has attempted to clarify the issues that need to be considered in evaluating the choices made by the New Zealand system.

Its conclusions are as follows:

♦ First, as long as the informal elements of performance measurement are not neglected, formal contracting and performance measurement probably improve performance.

♦ Second, although a government does need to consider both its ownership and its purchase interest in its departments, it generally shouldn’t take a traditional profit-maximising approach to ownership monitoring and shouldn’t operationally separate purchase and ownership monitoring.

♦ Third, in most cases it is better to hold organisations accountable for things toward the output end of the output-outcome spectrum; in some case, however, the motivational advantages of outcome-like contracts may outweigh their disadvantages.

♦ Finally, governments have a strong interest in employing systems that provide for routine and rigorous evaluation of the effects of outputs on outcomes, and the New Zealand system would benefit from improvements in this domain.
NOTES

1. This paper has benefited from comments by many people, including David Shand, Ian Ball, Jim Brumby, Charlotte Williams, Michael Wabster, Alex Matheson, and Maitland MacFarlan, and from discussions of the subject with many others, including Doug Andrew, David Ehrardt, Peter Gorringe, and Paul DiMaggio. However, the views expressed are those of the author alone. The paper was written while the author spent a summer internship at the Public Management Service of the OECD in 1994.

2. See New Zealand Treasury (1987), Scott, Bushnell and Sallee (1990), and Scott and Gorringe (1989) for statements of rationales for the reforms, and Hood (1991) and Pallott (1991b, p. 189) for comparisons of these of New Zealand’s reforms with others.

3. This point is due to Paul DiMaggio, Princeton University (personal correspondence).

4. Heymann (1988) contains a perspective discussion of how the need for formal controls varies with cultural norms (about Loyalty and corruption, for example) and with the personal goals of managers. Although his interest is in the need for input controls in government agencies, the problem is similar.

5. More precisely, we can describe the goal as maximising the expected discounted net value for the department’s services, where the net value of the department’s services is the sum of the consumer and producer surpluses resulting from their provision.

6. In the jargon of economics, any increase in producer surplus is likely to come at the expense of consumer surplus.

7. In principle, the government might use statistical techniques, such as multivariate regression to control for other influences on outcomes and thus isolate the effect of the agency. Although this sort of performance measurement has been used, it may often be impractically difficult or costly to use for routine accountability purposes.
REFERENCES


OSBORNE, David and Ted GAEBLER (1992), Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts


Chapter 2

AUTONOMY, RESPONSIBILITY AND CONTROL
THE CASE OF CENTRAL GOVERNMENT AGENCIES IN SWEDEN

by

Yvonne Fortin

Fondation nationale des sciences politiques, France

Introduction

The present central government management reforms in Sweden are, for the most part, the result of a process of change which in some areas started about fifteen years ago. For example budget flexibility was gradually introduced before 1988 when it was decided, among other things, to opt for a three year budgetary cycle. In human resources management the basis for change was a government bill on personnel policies passed by Parliament in 1985.

The background of the reforms of the last ten years can be described as follows:

♦ an increasing budget deficit and difficulties in financing the public sector;

♦ expansion of government functions accompanied by a loss of steering and control at all levels leading to a gap between goals and results achieved;

♦ public demands for increased service quality;

♦ increasing competition, reflecting technological developments and private sector demands for access to areas of public monopoly;

♦ increasing pressure for a performance-oriented public service organisational culture.

As in many OECD Member countries the Swedish reforms aim at ensuring that central government organisations in general and agencies in particular make the best possible use of resources. In short they aim at turning already autonomous but previously input-oriented organisations into result-oriented ones. Measuring outputs on a regular basis is now required of politicians and managers. Clear objectives have to be set, performance measures and indicators defined, results reported, discussed and used as an input to allocation of resources and to policy-making.
Agency heads are now accountable for achieving objectives, not just for following financial and administrative rules and procedures. In addition budgetary rules have been amended so that top managers now have increased authority to manage the resources allocated to their organisation.

The ongoing decentralisation leads to increased responsibility and impacts on accountability: management by results implies that those to whom managers are accountable have the power and the means of rewarding success as well as penalising failure.

The terms and conditions of appointment of agency heads and deputy heads are being reconsidered so that they fit the new framework of accountable management.

The characteristics of the central government system in Sweden from a managerial perspective can be summarised up as follows.

In Sweden there is a basic division of responsibility between the Cabinet and ministries on the one hand, and agencies on the other. Agencies traditionally have a very high degree of autonomy and decision-making power. Neither Cabinet nor Ministers may intervene in decisions of agencies where the exercise of authority and application of law are concerned.

Formal power for all governmental decisions rests with the Cabinet which is collectively responsible for them. The Cabinet makes decisions about important issues of governmental policy, while routine matters are in practice decided by ministers individually.

Agencies are accountable to the Government and the Government is accountable to the Parliament. Thus, there is no direct formal line of accountability between agencies and Parliament. The annual reports of agencies are not formally submitted to Parliament although they are publicly available as generally are all Swedish Government documents. Parliament receives information concerning agencies results in the Budget Bill and has in recent years pressed the Government to present more useful information on the results of public program. In addition to the budget process, the standing committees in the Parliament may question any minister or Director-General about agency activities and results.

In relation to accountability, there is another important feature of Swedish public administration, the purpose of which is to afford insight into and control over, public agencies. This is the principle of public access, whereby everyone domiciled in Sweden is entitled to study an agency’s files. An agency is not entitled to ask the reason why a person wishes to see a public file, nor to find out that person’s identity. For employees, it means that they are legally entitled -- with certain exceptions -- to pass information on to the mass media for publication. The agency is prohibited from investigating who has given the information to the media. These are the basic principles, although in practice there are of course a number of restrictions.

The Ministry of Finance is responsible for economic policy and the budget. In addition it has overall responsibility for the development of public administration. Agencies under the Ministry include the Agency for Administrative Development and the National Audit Office, both having responsibilities for public management reforms at the agency level, but from different points of view. The Ministry of Finance is also the central personnel organisation. The National Agency for Government Employers is responsible for collective bargaining and for developing and implementing personnel policies.
Currently there are 13 ministries. “Line” ministries are small and few have a staff of more than 200. Their main task is to assist ministers in policy making by preparing government bills for Parliament covering budget appropriations and other legal aspects of operations, giving advice on the need for reform programs usually prepared by special Government commissions, preparing and coordinating Sweden’s initiatives and positions within European Unions, controlling and monitoring the agencies and deciding on a decreasing number of administrative matters.

Each agency is assigned to one specific division, headed by an Assistant under-secretary in the relevant line ministry. Thus an agency has one contact with whom to discuss all substantive issues that may arise. In practice this does not exclude an agency-head from having informal direct contacts with the under-secretary or the Minister especially on policy matters.

The Assistant Secretary heading a division is responsible for the dialogue with the Budget Office of the Ministry of Finance Concerning the budget for each agency. Thus, the agencies do not deal directly with the Ministry of Finance.

There are about 330 agencies. They form part of the civil service which employs a total of about 234,000 staff, including public service corporations, line ministries and the national defence. Agencies contribute to policy making in that they are expected to submit proposals to government regarding the policies to be followed by them. Decisions concerning the implementation of policies are taken by them, not by the parent ministry. Agencies vary greatly in size, organisation, and structure. Individual agencies are responsible for their own accounting and financial management. The government sets the basic rules, in the form of financial and administrative ordinances. The National Audit Office is often given the responsibility to supplement these ordinances with more detailed instructions and advice on how to interpret the ordinances.

Agencies are headed by a Director-General appointed by the government, usually for six years. The Director-General is often assisted by a board whose members are appointed by the government. Lay board members come from different walks of life such as Parliament, the professions, client organisations, staff organisations and private companies. The board decides on the agency’s three years budgetary request, on its annual report and on regulations issued by the agency affecting the citizen. Some agencies have a board with only an advisory function. In these cases the Director General is responsible for the decisions mentioned.

The management team is composed of the Director-General and his or her deputies who are appointed by the Director General, from inside as well as from outside the agency. In Sweden there is no public service career system as in most other OECD countries, but rather an open appointment system. This means that every post is publicly advertised and that anybody is free to apply. The most suitable applicant according to skills and experience will then be appointed.

Below the Deputy Director-General level, each agency manages its own recruitment and appointment of staff.

The National Audit Office plays an active role in developing performance measurement and management. It has three main roles: financial auditing, performance auditing (effectiveness) and financial management (accounting, developing management systems, setting performance standards). It has considerable independence from the government and is fully independent in its auditing roles. It is headed by the Auditor-General who is appointed by the government for a six-year renewable term. It has a total staff of about 500.
In addition the parliament has its own auditors -- the Parliamentary Auditors. These auditors are members of Parliament and they have a small professional staff consisting of about fifteen performance auditors. The Parliamentary Auditors have a mandate to audit every aspect of state financed activities. They do so from a value-for-money or performance perspective.

The National Agency for Administrative Development (SAFAD) assists the Government and the ministries to review and increase effectiveness in state-financed activities. SAFAD carries out Government assignments and provides the ministries with analyses on such matters as organisational structures, the implementation of performance management, the division of responsibility between central and local government, the use of information technology and the development of an open electronic infrastructure for the public sector. SAFAD has a staff of total 100.

**Autonomy**

*Traditional autonomy, protected by law*

A characteristic of Sweden’s public management is the organisational split between the Government ministries and the agencies. The autonomy of Swedish agencies did not start with recent management reform. The central government in Sweden has been organised according to this principle since the beginning of the last century. The organisational division between ministries and agencies may be explained as a desire to distinguish between policy and implementation. Policy is the responsibility of the Parliament and the Government, while the implementation of programs is the agencies’ task.

Two matters of a constitutional nature are important in relation to how the Government controls the agencies. The first is that the agencies are subordinate to the Government as a collective body. Thus, an agency does not formally come under its Cabinet minister or ministry. This is because all Government decisions are taken by the Government as a collective body, not by individual ministers.

The second matter concerns the Government’s scope for giving directives to the agencies. The basic principle of Government control of the agencies is restricted by one important exception: the agencies themselves must decide on specific issues where the exercise of authority and application of law is concerned. On such matters, it is constitutionally prohibited for the Government to influence the agency’s decisions. However in today’s complex society, agencies have many tasks that do not involve the application of law. All agency activities not involving the application of law may be controlled by the Government as it wishes.

The Swedish example shows that autonomy itself does not necessarily lead to a management system which focuses on results. On the contrary it shows that managerial autonomy can co-exist with a system which is input oriented and where accountability focuses on adherence to rules and procedures.

**Content of autonomy**

Government agencies are wholly responsible for their own staff policies, accounting and financial management and to a large extent, their own investment decisions. They are few restrictions on the right of agencies to deploy funds for their own administrative purposes. To a large extent an agency is free to determine its internal organisational structure, select its working methods and combine production inputs in ways which best suit its activities.
Budgeting

Parliament allocates funds to agencies and government enacts directives to the agencies in a “government approval document”. This document has become more important in the management of agencies. It includes directives concerning agency objectives, performance requirements, resources and framework conditions. Agencies also are authorised to disburse funds in the form of transfers and investments.

The Government itself manages mainly funds which have been allocated to meet the administrative costs of the Cabinet Office and the ministries.

An additional feature of the budget process is the introduction in 1993 of a so called “frame appropriation” allowing 3 to 5 per cent of annual savings or credit. Agencies can carry over surplus funds from one fiscal year to the next. Agencies may overspend their current budgets but this is automatically deducted from the following year’s budget. The purpose is to grant more flexibility in the management of administrative costs and to reduce the “end-of-the-year effect” i.e. spending the remaining budget in order to avoid a reduction in future appropriations.

The agencies may borrow from the National Debt Office for investments in fixed assets (e.g. computers). The loan is provided on market terms and with a duration corresponding to the economic life of the asset. The interest expenses are deducted from the agency’s frame appropriation.

Fee-Based Activities

An important element of financial management in agencies is the rules concerning fee-based activities. Some agencies, through formal agreement by the Government, are able to use revenue from fees to cover operating costs. The Government decides on a case by case basis which agencies or areas within an agency will have a fee-based activity. The following principles serve as the basis for the management of fee-based activities:

* Full cost recovery is the primary financial objective. Thus the agencies may not charge excessive fees to make profits.
* The proposed level of the fees must be submitted by agencies to the NAO. NAO makes a formal recommendation on the level of the fees. The agency concerned then decides the level of the fees, in most cases following the recommendation of the NAO.
* Fee-based activities have to be reported separately in the annual report for each agency.

Human Resources Management

Human resources management is highly decentralised compared to most OECD Member countries. Yet decentralisation is still increasing.

Human resources management also operates according to the co-determination principle under which certain matters affecting staff members must be negotiated or discussed with the trade unions.
The general framework and objectives for human resource management are set at central level. Collective agreements concerning for example long term commitments from the State as an employer on such issues as job security and pensions are negotiated at central level. Collective agreements on salaries are negotiated at central level. These agreements determine the general percentage, or sometimes a lump sum that every civil servant will get and secondly the percentage that each agency will have the freedom to allocate to its employees, according to agreements reached after collective bargaining with the unions at agency level. This second component is increasing in importance.

The government has recently taken a new step which aims at increasing the influence of individual agencies on pay determination. In order to further increase the agencies’ powers as employers further, a new employer’s organisation was set up in the state sector: The National Agency for Government Employers. The purpose of the reform is to create an employers’ organisation that is more autonomous in relation to the Government. This is reflected in the fact that the new agency’s board and director-general are no longer appointed by the Government, but by a council of agency representatives. What is more, it means that neither Parliament nor the Government will henceforth have to approve the pay agreements concluded in the state sector.

In most agencies there is a central unit in charge of personnel matters. But this unit mainly deals with policy and provides support to managers. Agency line managers have full responsibility for their own decisions on definition of tasks and appointment and allocation of staff.

There is no fixed grading system. Such a system disappeared some twenty years ago. Managers have the power to decide on grades. The present grading system is task- or job-based.

Budget Management and Annual Reporting

Based on new rules issued on July 1, 1993, the annual report can be said to have taken the place of the budget request as the most important document in the budget process. The annual report consists of a performance report, an income statement and balance sheet and an appropriation account and financial analysis. Accrual accounting is used to measure and report costs. The performance report covers inputs, outputs, productivity and “measurable effects” or outcomes. The annual reports are scrutinised by the National Audit Office and are submitted to the Government for use in the budgetary process.

This system included, at an early stage of a reform, a 3-year budget cycle. An in-depth budget request is submitted September 1, every third year and includes performance data for the previous five years; proposed directions for the next three years; a three-year resource budget; proposed measures and objectives for results-based management; and a budget request for the first year of the three year period. Simplified budget requests applied in the other two years of the cycle. The objective of the system was to further devolve the budget authority and responsibility to the spending departments within a strategic management framework.

There has been a shift from this cycle to a more continuous flow of information. When the reform was first launched about half of the scheduled three-year budget allocations were deferred due to lack of adequate information. The economic recession further undermined the ability to make three-year commitments. Agencies are now subject to in-depth review at least every six years. The Cabinet Office and ministries now decide on the basis of what the activity requires when the in-depth examination will occur.
There have also been difficulties in implementing “agency-specific directives” which were supposed to advise the agency one year in advance on what information to provide for the in-depth review. Ministries were seen to be too occupied with immediate budget issues to focus on the needs for next year’s budget. There were also difficulties in agencies, parent ministries and the Ministry of Finance identifying and agreeing on strategic issues to be addressed and the process was found to be more time consuming than envisaged.

Ministries still have difficulties in formulating operational goals in such a way as to facilitate for the agencies follow up and assessment of their activities. The budget dialogue has improved although some difficulties remain. The link between results analysis and the proposals put forward on operations and resource needs is often weak. The in-depth assessments have not always been effective since the assumptions and financial conditions may have often changed during the period, reducing the relevance of the initial documents.

But overall the results are considered positive. Agencies are improving their competence in management by results and financial administration. The new budget process has resulted in improved annual reports which show more clearly what has been achieved and where performance can be improved. In the future development of the budget process performance management will be clarified to provide a clearer focus on analysing results as part of the agency’s detailed funding request. The Government is working towards further improvements in results information and its use.

**Accountability for results**

The Director-General is accountable for the results of the agency. This is clearly stated in the General Ordinance for Agencies.

The Director-General must ensure that the agency’s activities are conducted efficiently and effectively in accordance with relevant statutory provisions. It is his or her responsibility to ensure that citizens are provided with services of good quality.

It is the task of the Director-General to ensure that the staff is well aware of the objectives set by the Government and to systematically develop the staff’s competence.

The Ordinance states that it is the responsibility of the Director-General to follow up and review the activities of the agency. He or she is responsible for providing the Government with the analysis and information needed for the Government to decide the role and objectives for the agency.

**Prerequisites to Accountability**

Some prerequisites exist for a manager to be held accountable. First, he or she can only be held accountable for results achieved on the basis of management decisions which are clearly his or her own. Second, a balance has to be struck between stability and flexibility in the deployment of managers. On the one hand the manager must hold the position for a particular length of time so that at the end of a specific management cycle results can be attributed to his or her management, and he or she can be rewarded or penalised according to the results. On the other hand flexibility is important so that the employer (in this case the government) is in a position to remove a manager or appoint him or her to another job.
In addition performance must be properly measured. This is only possible if objectives set have been clearly stated and if relevant performance measures and indicators have been developed. The person or the organisation responsible for setting objectives and/or analysing performance reports has to have the knowledge and skills to do so.

The Director-General’s tenure

As mentioned previously the Director-General is appointed by the Government. The appointment is proposed jointly by the line ministry and the Ministry of Finance. The former looks at the candidates, their experience and qualifications and then discusses the level of remuneration with the Ministry of Finance. Greater attention is now paid to leadership and management skills.

Directors-General enjoy stability, being appointed for a six-year renewable period. From this point of view, results can be clearly attributed to them. Thus they can be held accountable for the outputs and some of the outcomes achieved by the agency, over the three-year budget cycle period.

However, designed for a system where organisations were not run according to management by results, this six-year security of tenure did not fit the new management system. Sanctions for poor performance were inadequate. A commission reviewing the Public Employment Act proposed to the Government some changes concerning the appointment of Directors-General in order to give the government more flexibility over the management of Directors-General. The Directors-General cannot be dismissed before their contract expires but the government can, for organisational reasons or if it is necessary “for the growth of the agency”, transfer the head of an agency to another civil service post. The Government has used this power on some occasions.

Decision-making powers of Directors-General

Another difficulty arose from the fact that some decisions at agency level were taken not by the Director-General but by the lay board. As mentioned above members of the lay board were empowered to decide on three important matters the agency’s three year budgetary request, the annual report and on an agency’s regulations affecting the public. This led to a dilution of the Director-General’s responsibility and accountability.

A committee was therefore appointed by the government to look at the role of lay boards. One point at issue was to increase and clarify the responsibility of Directors-General. The government proposition was criticised by Parliament and no fundamental change Concerning the role of the lay boards has so far been implemented. The Directors-General were however given the authority to select and appoint the deputy directors-general.

The Co-determination Act under which the Director-General must negotiate with the unions on any matter affecting staff members (e.g. the budget, recruitment, equipment, etc.) is also under review. The purpose is not to abandon the co-determination principle but to modify it by emphasising direct contacts and discussions with staff members at unit level (as opposed to formal high-level consultations with unions), consistent with decentralisation within agencies.
Setting objectives for agencies

It is recognised that Government objectives are often not set with enough clarity to enable agencies to develop their strategies and to measure performance. The Ministry of Finance is working to improve this state of affairs by implementing an educational program for the relevant civil servants in the line-ministries.

Decentralisation within agencies

The Swedish government is well aware that there should not be one set of rules at one level of the organisation and another for lower ones, and that performance-based management should operate at all levels throughout an agency. Many agencies have decentralised their management according to the principles of the new management approach.

But as in many OECD Member countries this has not always happened and further progress is needed. Result-based management must operate along the entire chain of command, from the drawing up of agency objectives by the government down to internal management at agency level, with clearly defined result requirements set for each part of the agency. There are good grounds at the present time for introducing tougher targets here as well as tougher requirements with regard to methods of developing targets.

The degree of decentralisation within agencies also has an impact on the quality of information in annual reports. Unless an agency internally makes use of the information and has an adequate information system, the information will be of poor quality when it is aggregated and passed on to the parent ministry.

Conclusion

Sweden, like other OECD Member countries, can be described as being at an early stage in the sense that the current reforms have been implemented for about five years only. However the government has already put considerable effort into identifying progress and weaknesses. The diagnosis is clear and comprehensive.

A particular feature of the current approach is the role of annual (performance) reports in informing the budget process. Yearly assessments of agencies’ outputs, as part of the budgetary process provides the government with an adequate information tool to help it understand on a regular basis what is happening and why, in addition to other studies or enquiries conducted by other central agencies.

After a few years the three-year cycle was in practice changed to a more flexible system. At present the Ministries decide on a year-to-year basis which agencies are to do an in-depth analysis. The main rule is that each agency shall carry out such a scrutiny at least once during a six-year period.

The implementation of management by results was accompanied by a decentralisation of financial responsibilities to the agencies. This decentralisation has been successful; more effective cash-management and a more rational use of state appropriations for running costs have been achieved.
From a comparative perspective Sweden seems to be unique in that its government system operates around two kinds of “responsibility centres”; the government as a whole on the one hand and individual agencies on the other. The combination of constitutional rules and management reforms make the whole government, as opposed to individual ministers, responsible for setting objectives, allocating resources to individual agencies, appointing and managing agencies’ heads. The Swedish system concentrates in one single body, the Cabinet, the decision-making powers to control an agency. By doing so it clarifies and increases government political accountability to Parliament.

In addition the Swedish system of government based on the extensive autonomy of agencies creates clear responsibility centres at this level too. Agency heads are responsible for managing the resources and are accountable for results to the Government.

It will be of great interest to watch the way responsibility and accountability will combine in practice, from a management perspective. From a comparative perspective it would also be interesting to discuss the positive and negative -- if any -- effects of having largely autonomous units before introducing the present kind of management reforms.
NOTE

1. I am grateful to Lars Dahlberg of the Swedish Agency for Administrative Modernisation, to Ulf Bengtsson, of the Budget Office in the Swedish Ministry of Finance, and to David Shand of the Public Management Service, for their comments and assistance.
REFERENCES


Ministry of Finance, *Annual Performance Accounting and Auditing in Sweden*.

Ministry of Finance, *The Public Sector Labour Market in Sweden*.


Introduction

There has been significant public sector reform in Australia in the 1980s and 1990s. This has occurred at all levels of Government - Commonwealth (federal), State and local government. The objective of public sector reform at Commonwealth level has been to shift the focus in Australia from inputs and compliance to outcomes.

This paper will consider benchmarking in an outcomes-focused environment and some of the benefits and pitfalls of benchmarking based on Australia's experience largely at the Commonwealth level. This focus should not be taken to suggest that benchmarking is not being used at State and local government level. In fact, benchmarking has been used extensively at those levels of government. The focus on Commonwealth experience is for reasons of time, space and detailed knowledge.

The first issue addressed in this paper is that benchmarking has different meanings. Managers sometimes understand benchmarking as a technique close to "process mapping", i.e. analysing work processes and improving them by comparing them with best practice work processes in other organisations. This could be called "micro benchmarking". Other managers understand benchmarking as comparing the results of similar organisations in order to better understand the performance of the organisations. This could be called "macro benchmarking". Can these two models of benchmarking be used simultaneously and how do they contribute to each other?

The second issue this paper will address is the challenges raised by benchmarking. Whether the focus is on benchmarking processes or on benchmarking results or outcomes, the problem remains the same, i.e. how to find appropriate criteria on the basis of which an activity can be assessed and improved. What are the "right" benchmarks? How do we find them? How do we avoid goal displacement in choosing indicators and standards?

The paper will also look at whether benchmarking can be used as a tool for evaluation (i.e. an independent assessment of activities) or as a tool for continuous improvement used and owned by staff? Are both aims compatible?
This paper is structured in three parts:

- conceptual issues;
- the lessons: factors of success of benchmarking;
- conclusions, drawn from Australian experience with benchmarking;

The Annexes outline major benchmarking projects in Australia.

**Part 1 — Conceptual Framework**

**What is Benchmarking?**

The most commonly used definition of benchmarking is the search to find and implement best practice. That search is for a better — preferably best — performing organisation undertaking similar business to one's own organisation, understanding how that organisation is able to achieve that performance and then incorporating those ideas into one's own organisation. As a technique benchmarking was pioneered in the private sector, one of the most well known examples being Xerox. In this paper, this benchmarking technique is referred to as process benchmarking.

This is a deceptively simple definition, which masks the complexities which can arise once benchmarking is put into place. Some of these issues will be addressed later in this paper. However benchmarking has a number of other meanings. These include:

- "results benchmarking"; and
- "benchmarks" or best practice standards.

The different concepts reflect different emphases rather than different techniques. In fact, there is a very close relationship between processes (process benchmarking), performance information (results benchmarking), and performance standards (benchmarks).

**Process Benchmarking**

Process benchmarking applies to the processes and activities that are used to turn inputs into outputs. Processes should not be considered to be formal rules, but rather the chain of activities or work practices.

By making comparisons of processes or developing process indicators, the primary focus of process benchmarking is to go behind the quantitative information provided by results benchmarking to establish the reasons for different levels of performance and incorporating best practice ideas into those processes and activities.

**Results Benchmarking**

Results benchmarking encompasses comparing actual performance of different organisations using performance indicators. They are not standards, (i.e. ideals or norms to be achieved), but rather
a suite of measures or proxies to determine efficiency and effectiveness. The focus is not on processes but more on results and outcomes.

For example, the comparative performance assessment of Commonwealth, State and Territory government trading enterprises (GTEs) is a results benchmarking exercise (see Annex 2). The purpose of the exercise was to develop a consistent set of indicators to allow the performance of GTEs to be monitored and to publish these on an annual basis. Comparisons are made between the organisations or jurisdictions rather than against a "benchmark", although the performance of a particular organisation could become the "benchmark".

**Benchmarks (ex ante standard setting)**

Benchmarks are a form of standard. Standards can be categorised into minimum, average and best practice standards. Minimum standards specify a level of service or performance that must always be achieved. Average standards imply that half of the service users will receive service which is below the average, and hence presentationally they have some difficulties. Benchmarks are best practice standards. Best practice standards have an aspirational character and should therefore motivate staff. Minimum standards, on the other hand, may provide little incentive to staff who may consider that better service could be provided.

The best practice standard can be at the process level — in this case the benchmark would be the processes of the best practice organisation. Best practice standards for processes may not be visible to clients and hence may require internal rather than external support to be attained. This is particularly true of support processes, such as accounts or personnel processes, which do not impact directly on external clients.

Benchmarks can also be targets for performance specifying a level of desired outcome or setting broader goals. They are often defined in the field of social activities, but could be extended to economic or fiscal targets. "Social" benchmarks are measurable standards which reflect the level and quality of services that citizens are entitled to expect. The benchmarks or standards may apply now, or be objectives for the future. The primary focus of social benchmarks is on outcomes — about what government programmes should achieve.

**Why Benchmarking?**

In the private sector, the primary rationale for benchmarking is the desire to maintain a competitive market position. In the public sector, many agencies are not in a competitive market position. Nevertheless, there are sound reasons for implementing benchmarking in the public sector.

**Benchmarking to Improve Performance**

Benchmarking is about comparison and comparison is a driver of performance. Benchmarking allows staff to find out for themselves how they are performing compared to others and can give them ideas that they can implement to improve performance. Benchmarking usually encompasses comparison with external parties. However, it is also possible to make comparisons between parts of the organisation which perform the same processes. Such internal benchmarking can generate a source of ideas — over time the practices in different regions may have diverged, leading to variations in performance in different regional offices.
The Australian Department of Social Security has undertaken extensive internal benchmarking. It has developed best practice models to assist regional offices to meet performance standards. In the area of processing applications for age pension, the national performance standard is for payment within 28 days after the claim is lodged. Internal benchmarking has contributed to an increase in the percentage of claims processed within this standard from 56 per cent in 1991-92 to 79 per cent in 1993-94. (As internal benchmarking was only one of the improvement initiatives put into place by the Department, it is not possible to establish a direct causal link between benchmarking and the improvement). Benchmarking the processes between the regional offices therefore involves comparison and can result in improved performance.

However the performance variation between different parts of the same organisation or even different organisations within the public sector carrying out similar processes, is likely to lead to incremental rather than major changes in efficiency. This may be partly due to similar cultures, and similar legislative environments. Internal comparisons may therefore lead to greater acceptance of the status quo. The greatest performance gains from process benchmarking are likely to come from comparisons with external providers. Given the major improvements in efficiency which have been associated with some contracting out initiatives (e.g. up to 70 per cent savings from the Commercial Support Programme in the Department of Defence), this suggests there can be particular value in external benchmarking with a private sector organisation.

Improving Performance Information

A further motive for benchmarking is to improve the quality of information on public services. The information gained can be used to meet accountability obligations and/or to demonstrate performance (for example in the context of the service being made contestable). It is vital that government has reliable ways of assessing the relative performance of programmes in order to be able to set overall priorities and strategies. Benchmarking in this context can also be used by one level of government to set targets for other subordinate agencies or specify required levels of service. Benchmarking can also be used in resource allocation decisions.

There is a danger that managers may not take an integrated approach to performance measurement and may perceive the different sorts of performance information to be an add on to their work, rather than being integral to their work. Process benchmarking, results benchmarking and social benchmarks are all part of the performance information picture necessary to determine how public sector organisations are performing at all levels.

Implementing Benchmarking

Some of the steps involved in implementing benchmarking are:

- understanding existing processes;
- selecting indicators;
- choosing benchmarking partners;
relating process benchmarking and outcomes;

creating links between benchmarking, continuous improvement and evaluation.

**Understanding Existing Processes**

Organisations often incorrectly assume that internal processes are well understood and documented. To develop new ways of working, it is crucial to understand how the current process works so as to know what is driving differences in performance. Understanding current processes usually requires some form of process mapping — a diagrammatic representation of the activities in the process, and how those activities relate to each other. Internal benchmarking is one way to develop this understanding, and is perhaps the most appropriate use of this form of process benchmarking.

**Careful Selection of Indicators**

The selection of performance indicators is one of the most crucial aspects of benchmarking. The quality of benchmarking is dependent on the quality of the performance indicators developed and the criteria for determining how performance is assessed against them.

Performance indicators have the potential to distort organisational behaviour. For example, when Australia Post first developed an indicator to measure performance of its mail sorting centres, it chose a simple, easily quantifiable, measure of output — the number of letters sorted in a day. A perverse consequence was that in some instances, pre-sorted mail was sorted again to boost apparent performance. Mail delivery was consequently slower. The performance indicator now used is the percentage of mail delivered on time from mail box to addressee, timeliness of mail delivery being important to clients. Linking indicators to what is important to clients is one way to reduce the potential for performance indicators to distort behaviour.

Assessment of performance cannot be determined against one indicator. A whole suite of indicators is necessary to give a complete picture of performance. In assessing performance against this suite any interrelationships between the indicators needs to be taken into account. In judging performance readers of the data will also need to consider the relative weighting of the indicators. For example, in relation to the indicators developed for hospitals as part of the Commonwealth/State service provision review (see Annex 3), if a hospital is performing well against the indicators for success of surgery but poorly against the access indicators, is that good or bad? What is more important? What do service users consider to be more important?

The presentation of performance information, particularly where comparisons are being made, requires a considerable degree of sophistication from the user, including an understanding of the policy environment and the factors which lead to differences in performance.

**Choosing Benchmarking Partners**

Determining best practice standards for processes is synonymous with finding a best practice partner — an organisation which illustrates best practice for all aspects of the process. However no one organisation is likely to illustrate best practice for each performance indicator. For example, a recent project examining personnel practices in Australia (see Annex 5) selected a best practice
organisation in the private sector for comparison with a number of public sector organisations. The best practice organisation was judged to have effective, efficient and fair human resources management practices which are an integral part of its overall business success. The criteria for selecting the best practice organisation included that it was balanced across all of its human resource activities, that it was not best practice in one area and worst in another. Importantly, the best practice organisation did not have the lowest costs.

Sources of information for selection of best practice partners include professional associations, best practice programmes, benchmarking user groups and consultants.

Both Process and Results Benchmarking are Necessary

A focus on outcomes does not mean that questions of process are unimportant — quite the contrary. There is an expectation of and accountability for good processes which recognises both that the end does not always justify the means (particularly in public management) and that good outcomes can arise from good processes.

As well as improving processes, process benchmarking can help clarify objectives and outcomes, particularly where benchmarking partners have such a different way of approaching similar processes that it causes the organisation to question those objectives. For example, the benchmarking of personnel services between government departments and private sector organisations gives rise to questions about whether the legal constraints or the rules imposed on public personnel services are really justified by the principles of public service.

By improving the quality of an organisation's outputs process benchmarking has a direct bearing on outcomes and hence on the clients of programmes and services. For example, service quality standards related to processing times or simplification of procedures may produce a better outcome for clients.

Results benchmarking should not be considered as the opposite to process benchmarking but rather as complementary. Comparing results across agencies and jurisdictions naturally provides the impetus for organisations to question why differences occur. For example, what factors are there that lead the better performers to achieve that level of performance? Are they environmental factors (e.g. legislative requirements, efficiencies due to size, volume) or do they arise because of the way in which the organisations deliver the service, i.e. because of differing processes?

Results benchmarking also raises questions about how the gaps in performance can be narrowed. One of the ways is through process benchmarking — understanding the processes that different organisations use, what the differences are and how the different practices can be incorporated. However, where the performance information is at a highly aggregated level (e.g. relating to a particular sector), it can be more difficult to translate this into process can be improvements.

Process and results benchmarking are therefore complementary, and should be used in conjunction with one another — the outcome of one form of benchmarking can inform the other.

Relate Process Benchmarking to Expected Outcomes

Process benchmarking has some potential shortcomings which need to be overcome, including the risk that managers undertaking benchmarking will become embroiled in detail and focus on
activities in isolation from whether they add value or increase the performance of the organisation. In the final analysis it is only possible to say that a particular process represents "best practice" if it leads to better results.

Within a results and evaluation culture, process benchmarking should be related to assessing client characteristics and to setting or realigning objectives. This does not mean that process benchmarking can only take place once client characteristics and objectives have been determined. Rather it means that process benchmarking should at some stage be informed by consideration of these issues to ensure that processes and objectives align (See Figure 1). In this sense, process benchmarking primarily has a reengineering focus — to ensure that the processes align with the outcomes for the programme or organisation and add value to achieving those outcomes.

![Figure 1](image)

Benchmarking in this context is similar to the concept of Total Quality Management. Once the desired outcomes have been established, the benchmarking analysis can determine the processes needed to deliver those outcomes. In particular, benchmarking allows the organisation to determine what parts of the process add value to the achievement of the outcome. By aligning processes with outcomes, there is a greater chance that the outcomes will be achieved.

Process and results benchmarking are complementary in these circumstances. Once the objectives have been set, it is then possible to determine the measures (or indicators) of success. Once such performance information of measures of success are available one can look at how other organisations are able to achieve better results in terms of efficiency and effectiveness. This involves process benchmarking then process redesign. Results benchmarking in this case is an incentive for process benchmarking.

Process benchmarking does not always occur in this way, including in Australia. The challenge is to better integrate process benchmarking within an outcomes culture. In particular, benchmarking should not be seen in isolation from questions about the purpose of the programme and organisation or whether there may be a more appropriate tool or technique to resolve particular issues.
An outcomes focus means a sharper focus on the objectives of programmes. It means focusing on how inputs and outputs contribute towards the achievement of those objectives, which include fair treatment of individuals, ethical behaviour and the like.

Performance information and benchmarking are crucial parts of an outcomes focus. Organisations need know how they are performing and that the processes that they use contribute to the achievement of their objectives. Performance information is one link in this chain. Other parts of that chain include a focus on client characteristics, a focus on targets and objectives, and evaluation.

There are flow ons from one part of the chain to the other. For example, a greater focus on client characteristics and requirements leads to better targeting and better defined objectives. More relevant and clearer targets and objectives assist in the development of better performance information. Better performance information improves knowledge of the operations and results of activities and assists the evaluation of outcomes. Better evaluation assists managers to focus on clients and improve performance against objectives.

Process benchmarking can in fact be evaluative, particularly where the process being benchmarked is part of the core business of the agency. Examining processes at the core of an organisation's business should include investigating client attitudes in order to measure the effectiveness (in terms of the impact on clients) of the process. Measurement techniques which go beyond a simple client satisfaction survey are likely to establish reasons for client attitudes, and have the potential to raise questions about outcomes. The Department of Finance's process benchmarking experience (see Annex 1) certainly had an evaluative flavour.
However, *benchmarking* also has a continuous improvement focus. Improvements do not only come from asking fundamental questions about objectives. Improvements to processes at a lower level can also produce significant efficiency gains.

If benchmarking is undertaken outside of the logic envisaged in Figure 2, the question which should guide managers is whether the results of the benchmarking will be worthwhile and cost effective. For example, a major benchmarking project which recently examined personnel practices in Australia, found that there were significant differences in performance between Australian Public Service organisations and a best practice private sector organisation. Many of these differences can be attributed to the different legislative environment. Had one of the organisations which participated in this study decided to undertake internal process benchmarking of personnel practices, or even external process benchmarking with other Australian Public Service organisations, they may have been able to make some performance improvements, but the impact of the legislative framework would not have been apparent. Even if the organisation had selected an external benchmarking partner, significant differences in performance which could be attributed to the legislative environment may well have been beyond the power and influence of a single organisation to change.

In other words, managers need to establish that the benefits in terms of continuous improvement of the process would outweigh the costs of undertaking benchmarking. In doing this, managers need to ensure that they make the most use of existing performance information. They need to understand the environment in which they are operating and to be clear on the issues they are addressing by benchmarking. They also need to ensure that benchmarking is in fact the most appropriate tool to address those issues.

**Part II — The Factors of Success and Potential Pitfalls**

From the Australian Government's experience with benchmarking, as outlined in the Annex to this paper, there are many factors of success for benchmarking and correspondingly pitfalls to be avoided. The factors of success for benchmarking are very similar to those for good evaluation. They include involving stakeholders, having well selected performance indicators and analysing reasons for differences in results.

**Involvement of All Stakeholders**

The comparative nature of benchmarking and the potential for changes to work processes and resource allocation decisions or relative priorities arising from benchmarking mean that it is crucial to involve all relevant stakeholders (clients, managers, employees and unions) in the selection of performance indicators and benchmarks and in the analysis of results against those indicators and benchmarks. The Department of Veterans' Affairs for example, initially found that unions were not in favour of benchmarking, and were resistant to comparisons with the private sector. By involving unions on the Steering Committee, they were able to undertake both internal benchmarking, and benchmarking with other organisations in the public sector.

Process benchmarking is not just a dispassionate examination of numbers. It involves determining what drives differences in performance and can involve making decisions about whether processes (i.e., how people do their jobs) are the most appropriate. Involvement of staff is therefore crucial.
Understanding The Organisation

Process benchmarking requires a thorough understanding of the processes used by the organisation. If comparisons are made without understanding internal processes it is difficult to determine what is driving any differences in performance. Internal benchmarking is a useful way of developing an understanding of internal processes.

Selecting Processes Which Are Important To Clients

One aspect of benchmarking is its potential to improve the quality of services provided to clients. The benchmarking literature suggests that the processes selected for benchmarking should be those which are of most importance to clients. These are the primary processes that deliver services to clients, rather than support processes such as human resources, which are less visible to clients. In addition, the performance indicators to measure the process should include some indicator of client satisfaction with the process and the output of the process. Benchmarking can therefore lead to a better understanding of what clients think about the process and the impact of the process on clients.

This is likely to involve more than merely asking clients whether or not they are satisfied with the process or service. It is important to establish what leads to clients' satisfaction or dissatisfaction.

As well as improving the processes which impact upon clients, benchmarking can also provide information on the standards of service that customers can expect.

Selecting Performance Indicators

The performance indicators selected for benchmarking need to be comprehensive in that, as a set, they present a full picture of performance of an organisation, programme or process. Regard also needs to be had to whether the performance indicators can result in distortions to organisational behaviour or goal displacement. A consultative process similar to that undertaken in the Commonwealth/State service provision review (See Annex 3) is one way in which relevant performance indicators can be developed.

Even with extensive consultation, the potential for goal displacement is unlikely to be eliminated. Ongoing review of indicators and replacement where goal displacement is evident is necessary.

Benchmarking Partners

Selecting benchmarking partners is often dealt with as a simple process in benchmarking training or benchmarking manuals — find someone who is performing better than your organisation and determine why. Benchmarking partners need not be obvious. For instance, a company making car parts might improve its distribution process by learning from a company making computers. A hospital might improve its admissions processes by copying those of best performing hotels. Such comparisons involved the comparison of similar processes, even though they take place in a very different environment.

However environmental factors are relevant, particularly in the public sector. The comparisons need to be valid. For example, the Department of Veterans' Affairs is unlikely to be able to use the
processes used by private sector debt collection agencies when collecting debts from elderly war veterans. A more appropriate partner would be a Health authority, seeking to collect debts from overseas visitors, where similar sensitivities (such as ill health) can apply.

One of the issues faced by public sector organisations seeking benchmarking partners is the lack of comparative information on the performance of other public sector organisations. This comparative information may be necessary to determine who would be a suitable benchmarking partner, particularly for process benchmarking. For example as part of the four agency benchmarking review (see Annex 4) participating agencies have had to conduct questionnaires to establish current levels of performance and hence establish which agency may be the best current performer.

**Recognising The Link Between Results And Process Benchmarking**

The best results from benchmarking are likely to come where the organisation recognises the links between process and results benchmarking. This is because results benchmarking does not provide the complete answer — determining what is driving differences in performance is as critical as establishing that those differences exist. Process benchmarking can help highlight where the differences are the result of processes and how those processes can be improved to lift performance. Similarly, process benchmarking cannot take place in isolation from consideration of the overall organisational performance, consideration of client characteristics and organisational objectives.

However, this is not straight forward. Process and results benchmarking operate at different levels — one focuses on process and the other on programmes or organisational objectives. For example, performance indicators for processes and results can be different. At the process level, performance indicators will relate to the efficiency and effectiveness of the process (how the programme is delivered) rather than the programme results.

For example, the performance indicators for the Department of Finance's benchmarking of the estimates process (see Annex 1) included timeliness (whether data was entered within the deadline) and elapsed time (time taken from commencement to completion of the process). These are indicators of efficiency of the process. Timeliness at the programme level (which has as its overall objective to facilitate the systematic development, implementation and public presentation of the Government's budgetary strategy) would relate to timeliness in the provision of advice to Government, rather than the input of data into a Departmental data base.

Without performance information at both levels it is therefore difficult to determine what is driving differences in performance. Results benchmarking, in particular, is the start of a questioning process designed to determine reasons for differences — possibly including undertaking process benchmarking to document and measure differences in processes.

**Finding a Balance Between Stability and Change**

Process and results benchmarking involve an element of target or standard setting. A significant issue is what happens once the standard has been achieved. Should ongoing performance be monitored against that standard or should a new standard be set? At the process level, what is best practice will evolve. Other organisations will emerge as the best practice models. This implies that process benchmarking should be a continuous process — that a process cannot be measured and reengineered only once. However, process benchmarking is not without costs — including the cost of the impact of change on staff. This suggests that a balance needs to be found between stability and
change. The same applies to results benchmarking. If standards are met and not changed, it can breed complacency. Constant change, on the other hand, can be dysfunctional.

Incentives

For the use of benchmarking to become widespread in the public sector, there needs to be incentives for managers to implement it.

The treatment of savings arising from benchmarking has the potential to be either an incentive or disincentive for managers to implement benchmarking. Managers need to know where they stand — whether the savings will be returned to the Budget, whether they can be retained by the agency or a combination of these approaches. A possible approach where the benchmarking is self-initiated is that the savings are shared between the agency and the Budget, and where the improvement technique is initiated centrally (e.g. by a central Budget agency), savings are returned to the Budget.

One of the potentially most powerful incentives for managers to implement benchmarking is contestability. Contestability is the potential for another supplier — either in the public sector or the private sector — to provide the service rather than the traditional source. It does not necessarily mean the transfer of provision of services to the private sector or contracting out. For example, the services could be delivered by a different arm and/or level of government — the possibility that competition could become a spur to innovation or the adoption of best practice by the traditional provider.

Benchmarking is a useful tool for managers looking to prove their performance in a contestable environment. Managers can use it defensively (i.e. to prove their current level of performance) and proactively (i.e. to search for new and better ways to deliver their services).

A Government Requirement To Benchmark?

Rather than rely on incentives for managers, it would be possible to have a Government decision requiring managers to implement benchmarking.

The potential success of such a requirement may depend on the nature of the benchmarking initiative. In Australia, for example, the benchmarking of Government Trading Enterprises (GTEs) has been successful in presenting better information on the performance of GTEs. Similarly, the benchmarking review of Commonwealth and State service provision is also likely to be a successful initiative. Part of its success has been the highly collaborative way in which the performance indicators have been defined and developed.

However, a requirement to implement process benchmarking has a number of difficulties. Process benchmarking is only one tool that managers can use to implement best practice into their organisation. A requirement for managers to implement benchmarking may well result in a "compliance response", without sufficient consideration being given to whether benchmarking is the correct tool for the particular issue to be addressed.
Benchmarking Requires a New Culture

A key success factor for benchmarking becoming integrated into organisational management are cultural factors. These cultural factors operate at two levels — at the organisational level and at the national level.

At the organisational level, benchmarking requires a culture where there is comfort with the notion of comparison — that our organisations are not the sole custodians of good ideas nor are so unique that they cannot compared with others. Given that the majority of public sector organisations do not operate within a competitive market environment, many public sector organisations are likely to be uncomfortable with this notion, particularly with making comparisons with the private sector.

At the national level, the culture needs to be one where benchmarking and performance information are used to understand rather than judge. If the information gained from benchmarking is used to make judgements without detailed understanding and analysis of the factors which lead to difference and how those differences may be resolved, it will be more difficult for benchmarking to become accepted.

Part III — Conclusions

One of the main benefits of benchmarking is that it allows organisations to develop a better understanding of the processes that they use to produce outputs, and the link between those processes and the outcomes that they intend to achieve.

The other main benefit is that it can be a driver of cultural change. It allows organisations to recognise that they are not unique and that there are many potential sources of ideas for performance improvement, including organisations which, prima facie, appear to be very different.

Both process benchmarking and results benchmarking are useful. Benchmarking of processes can lead to questions regarding the objectives of the organisation. For example, the Department of Finance process benchmarking exercise gave rise to questions about the appropriate role of the Department in the management of the estimates of government expenditure (see Annex 1).

The benchmarking of results invariably leads to questions about why differences in performance occur. Are they the result of factors such as the volume of business or are they due to differences in the processes that organisations use to produce outputs and which lead to different outcomes? In other words, results benchmarking and process benchmarking go hand in hand.

The links between process and results benchmarking mean that the starting point is not crucial. What is important in both process and results benchmarking is that organisations determine what is driving differences in performance. For process benchmarking, this means placing the benchmarking within the context of programme or organisational objectives. For results benchmarking this means looking beyond the data against indicators and examining the environmental context including the processes used by the different organisations to determine whether it is these or other factors such as policy or legislative differences which are driving differences in performance. In other words, it is important for organisations to recognise the links between process and results benchmarking, and that at some point the organisation use both techniques to determine what is truly driving differences in performance.
The other important conclusion from the Australian benchmarking experience is that there are strong links between benchmarking and evaluation. In Australia, these links are increasingly being recognised in the reviews of programmes and organisations commissioned in the context of the Commonwealth Budget. For example, a recent review of the Child Support Review Office (an agency which examines disputed assessments of child maintenance) evaluated the effectiveness of its review processes in meeting client needs and undertook process benchmarking to examine the efficiency of administrative procedures. This recognises that process benchmarking is more likely to be successful if it has regard to programme and organisational objectives.

Both benchmarking and evaluation are about processes and results. Very similar methodologies apply for successful benchmarking and evaluation — the involvement of stakeholders, the careful selection of performance indicators and close monitoring of the process used.
Annexes

MAJOR COMMONWEALTH BENCHMARKING PROJECTS - ANNEXES

Annex 1 — The Department of Finance's Benchmarking Initiative: Reengineering Processes Through Benchmarking

The Department of Finance is one of a number of agencies which have undertaken process benchmarking in the Commonwealth government sector. The project examined the estimates process, which underpins the Department's development of the annual Commonwealth Budget. The estimates process is part of the core business of the Department.

The project was an internal benchmarking exercise — comparisons were made between parts of the Department undertaking the estimates process rather than with an external provider. This allowed the Department to develop a better understanding of its current processes, gain some familiarity with benchmarking in practice and determine any methodological issues to be resolved prior to making comparisons with external organisations.

The development of performance indicators for the estimates process was an important part of the project. The indicators were developed in consultation with various stakeholders, although not with client Departments and agencies. Nevertheless "client satisfaction" was one of the performance indicators for the process. Of the indicators, only accuracy and reliability already existed — the others were developed for the purpose of the benchmarking project and data was collected against those indicators for the first time.

In developing the performance indicators, regard was had to the need to measure the efficiency and effectiveness of the process and the impact of the process on client Departments and agencies.

**Efficiency**
- Resources used per transaction
- Timeliness (whether process completed on time)
- Elapsed time (time between commencement and completion of the process)
- Impact (in terms of time and workload) of the process on clients

**Effectiveness**
- Accuracy and reliability of the estimates derived from the process
- Client satisfaction
- Quality of audit trail (i.e. data recording)

The performance of different sections of the Department was measured against these indicators, and the relationship between that performance and the processes used by each section was determined wherever possible. This proved to be difficult. Complexities included different relationships between Finance officers and officers in Departments and agencies and differences in the types of estimates managed. As a consequence no one area included in the study was considered to be the "benchmark".
Rather, general conclusions were made regarding how apparent differences in approach led to different performance, particularly in terms of efficiency.

The recommendations included those which could be categorised as continuous improvement in nature (e.g., relating to the timing and quality of information provided to departments and agencies as part of the Budget process) and more fundamental questions about the way in which the Department of Finance manages the estimates process.

One of the major recommendations was that the Department withdraw from the construction of estimates itself as a means of controlling the quality of estimates (in terms of accuracy and reliability). Instead, the report suggested that the Department devolve responsibility for estimates construction to agencies, and seek to control the quality of estimates by agreeing a methodology between Finance and the agency. This could be characterised as a quality assurance rather than a quality control approach. This a very far reaching issue which goes to the heart of the Department's role.

Successes and Pitfalls

Consultation with client Departments and agencies worked well. To measure client satisfaction a series of detailed interviews were held with client Departments and agencies, to find out what aspects of the process could work better from their perspective and why. These interviews lead to a much better understanding of the issues of concern to other Departments and agencies.

The benchmarking project has also lead to a better understanding of the estimates monitoring process. The processes used by various parts of the Department had not been previously documented and the Department did not have a good understanding of what parts of the process consumed the greatest resources. The Department now has a better picture of the different approaches to estimates monitoring, the costs and benefits of those approaches and the reasons behind those approaches. This information will enable the Department to ask whether the different approaches are appropriate.

The Department did not get everything right. The project was perceived by some as too judgmental. It is important to recognise that process benchmarking is not just dispassionate examination of data. It involves determining what drives differences in performance and can involve making decisions about whether processes (i.e., how people do their jobs) are the most appropriate. These decisions need to be as objective as possible, taking into account the performance information and relevant environmental factors and should ideally be made in close consultation with the staff concerned.

Another lesson from the exercise is that the cost of obtaining data from non-automated sources is very high. Other Commonwealth agencies have similar experiences. This in part reflects the fact public sector management information systems have traditionally focused on obtaining data to meet the requirements of financial statements and accountability obligations, and have generally not been focused at the process level. This is an issue that public sector organisations will need to take into account when determining whether process benchmarking will be cost effective and whether their organisation is ready for benchmarking.

Comparison with Others

While the internal process benchmarking project was necessary to understand the Department's process and the process benchmarking technique, the project highlighted the need for comparison with
external organisations. In particular, comparisons need to be made with organisations which have alternative estimates management approaches which may be able to be adopted by the Department in developing a quality assurance rather than quality control approach to estimates management. During October 1995, the Department will be conducting a survey of practice in certain OECD countries and in Australian States.

Annex 2 — Government Trading Enterprises (GTE) — Results Benchmarking to Improve Performance

The benchmarking of GTEs in Australia, covering both Commonwealth and State government organisations, is an exercise to develop a consistent set of indicators to allow the performance of GTEs to be monitored and compared. The focus is thus on results benchmarking.

The analysis of performance of GTEs is very important given their impact on the economy. GTEs currently contribute approximately 10 per cent of gross domestic product. They employ about 3 per cent of the Australian workforce and account for almost 14 per cent of gross fixed capital expenditure. They provide important services to the community and have an impact on the cost structure of industry. Many also operate as monopoly providers.

Three annual reports of GTE performance have now been published. The latest report, published in April this year, covered the years 1989-90 to 1993-94. This report is a comprehensive two volume document which looks at the performance of 59 of Australia's GTEs. The report includes:

- an analysis of the financial performance of GTEs by jurisdiction;
- the report cautions of the need for the comparative information to be viewed in context.

- commentaries on each industry;

◊ where relevant, these have been broken down into separate activities to ensure comparability. For example, in the electricity sector, electricity authorities engaged in distribution have been reported separately from those which provide other services or undertake both generation and distribution.

◊ an analysis of major changes to GTEs, such as industry restructuring, privatisation and commercialisation.

The range of performance indicators is extensive and covers financial performance, efficiency, effectiveness and service quality. The indicators of financial performance are identical for all GTEs regardless of sector. They include return on assets, return on equity, dividend payout ratio and debt to equity. The financial indicators give an indication of the return on the government's investment in the GTE, as well as the financial soundness of the GTE e.g. reliance on borrowings.

Efficiency and effectiveness indicators vary across sectors. Some examples drawn from the rail sector, where both the Commonwealth and the States operate GTEs include measures of efficiency including employee productivity (e.g. net freight tonne-kilometres per employee) and train kilometres per level crossing accident (effectiveness measure).
Data for the service quality indicators is less well developed than for financial performance. Where the data is available, the indicators tend to be industry-specific and cannot therefore be directly compared between GTE sectors. Examples of rail sector service quality indicators include service cancellations and on time running. A concentration on financial performance at the expense of quality indicators could cause distortions in organisational behaviour and potentially in Government decision making.

The performance information from the GTE benchmarking can be used on several levels.

GTEs tend to operate on a largely autonomous basis, with their Boards of management responsible for internal GTE management. The comparative performance information can be used by individual GTEs to determine how to improve aspects of their performance. For example GTE may find the best performing GTE within (or even outside) its sector, and use that organisation as a benchmarking partner for process benchmarking — to determine what is driving differences in performance. However, willingness to be a benchmarking partner could depend on whether the GTEs competitors.

The performance information on GTEs also feeds into strategic government decision making. The performance of Commonwealth GTEs is considered by the Structural Adjustment and Trade Committee of Cabinet. The information from the GTE exercise, together with information from GTE annual and interim reports and corporate plans is used to assess performance. Options open to the Government include restructure (e.g. capital injections), adjustment of dividend payments, changes to the management of the GTE or possible sale of the GTE. The appropriate option depends upon the level of the performance of the GTE, and what is driving that performance.

Annex 3 — Review of State and Commonwealth Service Provision — Results Benchmarking to Set Best Practice Standards

The Review of State and Commonwealth Service Provision was commissioned by the Council of Australian Governments (COAG) at its July 1993 meeting. (COAG is an annual meeting of Commonwealth and State heads of government). The purpose of the review is to develop performance indicators for areas of Commonwealth and State service provision and to collect and publish data against those indicators to allow governments to assess the relative efficiency and effectiveness of their service provision. The focus is on results benchmarking.

The Review is developing efficiency and effectiveness performance measures for:

- education - including schools and vocational education and training;
- public housing;
- community services - including the provision of crisis services and aged care;
- hospitals; and
- justice - including corrective services, courts and police.
Performance Indicators

The performance indicators selected are the result of considerable collaborative effort between the jurisdictions. This is a very positive process. It ensures that the policy environment of each jurisdiction is understood and taken into account in developing the indicators. It also ensures that there is commitment to the exercise and increases the chance for sound decision making based on the published data. Indicators for education and hospitals are illustrative of the performance indicators developed through this collaborative process.

Education

Efficiency indicators include expenditure per student, and staff/student ratios.

The most important effectiveness indicator for education is student learning outcomes. The working group will be trialing the use of teacher assessments, modified to ensure comparability, as a means of measuring learning outcomes. This will initially be undertaken in one key learning area, to determine the effectiveness of this approach. Other effectiveness indicators relate to retention rates, destination of school leavers and student attitudes to school.

A further effectiveness indicator which could be developed relates to the broader socialisation objectives of school education. This will not be included in the 1995 report. However the evaluation of the review will not only cover how data for existing indicators should be improved, but also whether the existing indicators are the right ones and measure the right things.

Hospitals

The hospitals working group has developed efficiency indicators and effectiveness indicators to measure both quality and access issues. Efficiency indicators include cost per case-mix separation, capital usage, labour costs per weighted separation, cost of treatment per outpatient and inpatient average length of stay.

Effectiveness indicators for quality relate largely to the success of surgery — the rates of emergency patient re-admissions; post-operative infections and unplanned return to the operating room. There is also a customer satisfaction indicator, which will be measured by surveys.

Effectiveness indicators for access largely relate to waiting times — for elective surgery, accident and emergency and out patients. They also include variations in intervention rate (e.g. caesarean, tonsillectomy, hysterectomy).

Comparability of Data

As well as determining the indicators, collaboration was necessary to define what is meant by each indicator. Without a clear and agreed definition of what each indicator meant and how performance should be measured against the indicator, there was a danger that published performance against the indicators would not be comparable.

Comparability is still an issue despite tight definition of the indicators and is influenced by differences between different governments. Some of these factors are:
• differences in size and geographic dispersion of the population;

• differences in mix of clients:

• varying public/private sector roles;

• variations in the types of services provided (i.e., policy differences); and

• data limitations.

To enhance the comparability between jurisdictions, users of the data will need to consider the whole suite of indicators to ensure that they have a complete picture of performance.

Initial Report

An initial report of the Commonwealth/State review is due to be published in December 1995. The report will outline the results of the exercise to date, including comparative data for all the efficiency and effectiveness indicators developed. The report will not set benchmark targets, but will present a comprehensive picture of performance in service delivery areas.

The report will contain the best information available at the time and will highlight areas for further work. Future reports will make on-going improvements to the quality, consistency and completeness of performance indicators and data. As well as addressing any gaps in data, relevant questions to be addressed in future years will be whether the current performance indicators are comprehensive and are the right ones, particularly from the point of view of the service users.

The potential use of the information from the review is similar to that for the GTE exercise. The different governments will be able to use the performance information to inform decision making, for example, in making resource allocation decisions.

Individual service deliverers will also be able to use the comparative information as a basis for changes to how they undertake their business-process re-engineering. The collaborative nature of the exercise and the provision of data by individual jurisdictions reduces the potential for the information to be used defensively. In determining why there are differences in performance, the service deliverers may well find that a factor is how services are delivered i.e., the results benchmarking may well drive process benchmarking.

The relative shares of Commonwealth taxation revenues allocated to different States is a matter of ongoing political controversy with claims that some states receive a disproportionate share being answered by claims of higher costs (because of greater size or population dispersion) or greater efficiency. This review will increase the factual basis of this discussion.

The issue for service deliverers and Governments to address is the link between results benchmarking information and processes. The driver of differences will not be immediately obvious — is it the processes or other environmental factors such as different objectives, differences in resourcing or different legislative requirements?
Annex 4 — Four Agency Benchmarking Review: Process Benchmarking

A major process benchmarking initiative at the Commonwealth level is known as the four agency benchmarking review. In the 1993-94 Budget, the Federal Government decided that four of the Commonwealth's large service delivery agencies [the Departments of Social Security (DSS), Veterans' Affairs (DVA), Employment Education and Training (DEET) and the Australian Customs Service] should participate in a benchmarking review. Each agency was asked to identify and then implement best practice in a range of processes associated with service delivery.

The three priority areas for the agencies were information technology, in particular applications software development and support processes, debt control, and claims and entitlements processing.

**Information Technology**

The focus on the Information Technology component of the benchmarking review has been on results benchmarking. The Australian Taxation Office has also been participating in this aspect of the review.

The benchmarking of the applications software process has centred on developing standards and procedures for recording effort on software development and cost attribution. An important part of the project has been to develop a method to ensure comparability of data on cost and effort. Software development projects differ in size, complexity and length of time for completion. The agencies participating in the review have decided to use function points (which measure complexity in terms of the number of screens viewed in a given application) to allow comparisons of cost and time.

Having a means of measuring complexity of program design will allow agencies to compare their performance with other organisations, including through the Australian Metrics Software Association, which maintains a database on applications software development.

The Australian Customs Service has found that the work undertaken to date has enabled their Information Technology area to identify areas of wasted effort and has assisted in estimating how much time is required to complete certain programs.

DEET has also undertaken external benchmarking with the banking sector and has drawn on the information from two external benchmarking partners to establish a usability laboratory for testing new applications. It intends to refine the operations of this laboratory to reduce the lead times for the testing of new applications.

**Debt Control**

Work on the initial development of the benchmarking model was through internal benchmarking by the DVA of its debt control for income support activities. This involved comparison between regional offices in each State. The report from this internal benchmarking exercise included recommendations for changes to work processes to ensure greater consistency across regional offices and the setting of a range of performance indicators, standards and targets to be used for future monitoring.

Agencies in the four agency benchmarking review are now drawing up plans for benchmarking with both public and private sector partners. This will involve the development of a questionnaire, to
be completed by each agency, which will be used to establish a cross-agency debt control profile. This profile will in turn be used to identify best practice and opportunities for improvement. That is results benchmarking will feed into benchmarking of processes.

DSS has also undertaken an internal process benchmarking study of its debt recovery arrangements. The measures undertaken during the study included data matching, process mapping and performance evaluation reporting. As a result of this benchmarking study and other management improvement initiatives, DSS was able to increase the recovery of overpayments by around A$100 million per annum.

**Claims/Accounts Processing**

This aspect of the benchmarking review will examine a process which has a direct impact on clients whereas most of the other benchmarking activities focus on processes which are essentially internal support processes to the core business of Departments. This part of the benchmarking review is still at an early stage.

All four agencies in the benchmarking review have agreed to participate in a survey of general claims processing to identify common processes and standards as a basis for benchmarking between the four agencies. In the case of DEET, they will also be developing a questionnaire specifically to analyse entitlements and claims for education and income support areas of the Department.

**Annex 5 — Achieving Cost Effective Personnel Practices: Results and Process Benchmarking Combined**

A further benchmarking exercise which has recently been completed is an examination of the cost of personnel practices in the Australian Public Service (APS). The project, Achieving Cost Effective Personnel Services, was commissioned by the Management Advisory Board (MAB) and the Management Improvement Advisory Committee (MIAC).  

The project collected information on personnel practices from a number of public sector agencies and private organisations and compared the practices for the public sector agencies against a best practice reference point — the performance of one of the private sector organisations. That organisation was selected because it was judged to have effective, efficient and fair human resources management practices, which are an integral part of its overall business success. The best practice organisation was used for all human resource areas, to avoid the potential for inconsistencies caused by using performance measures from organisations which may perform well in one service area at the expense of other service areas. Importantly, the best practice reference point was not the lowest cost.

The project used both results benchmarking and process benchmarking. The project collected global information for total personnel services and specific qualitative and quantitative information for particular personnel services (i.e. results benchmarking). The project also identified the processes and practices that significantly contribute to costs and effectiveness (process benchmarking). The aim of the project was to identify what was driving the differences in cost and effectiveness, including the extent to which this was attributable to the processes used.

Some of the findings of the review are that the direct costs of delivering human resources services in the agencies surveyed is more than twice that of best practice. It also found that APS...
practices are resource intensive in administrative and processing tasks compared to the effort put into more strategic human resource activities. For example, the processing of leave in the agencies surveyed costs A$20 million per annum, with almost 75 per cent of that cost being for leave of less than 3 days and nearly 60 per cent for less than one day. The report also found that the average Higher Duties Allowance (HDA) payment was A$70, with a processing cost of A$25.

In some instances, the differences in performance were attributable to different processes used — particularly where there were differences between organisations with the APS. However, many of the cost differences can be explained by legislative requirements. It could be argued that it is not appropriate or possible for the APS to implement arrangements based on private sector practice because of considerations such as equity issues.

While the APS should not automatically follow the best practice model because it is less costly, the project raises some significant questions. Are the significant differences in costs necessary? Is it possible to change personnel practices without impacting on equity issues? Are the complex rules and regulations which give rise to some of the additional costs still appropriate? Is the complexity consistent with an environment of a flexible and mobile workforce?

The report from the project will shortly be published and disseminated widely in the APS. Some of the issues raised in the report will be addressed on a service wide basis. These include reviews of the incidence of higher duties allowance and short term leave, with the intention of substantially reducing the number of personnel transactions in these areas and a review of permanent part-time work provisions.

There are also a number of pilot projects under way which aim to improve the personnel practices in certain agencies. These will concentrate on issues which can be addressed immediately, particularly the differences in processes used by different APS agencies. The pilot projects will also provide information on those policy and other issues outside the control of individual agencies which inhibit the achievement of best practice.
NOTES

1. An outcome is the real impact of a public activity as opposed to the quantifiable and immediate outputs. For example the output of a Department of Employment could be the number of unemployed people trained or finding a job whereas the outcome would be the reduction of unemployed or the role played by the department in helping to reduce unemployment.

2. The Commercial Support Programme is a programme to contract out non-combat activities (such as plant maintenance and catering). The programme has resulted in savings to the Commonwealth to date of more than A$300m; ongoing savings are about A$100m per annum. In-house bids have succeeded in roughly a third of cases, but at lower ongoing costs than the same team had previously incurred.

3. What this paper calls evaluation is not the simple measurement of results but the capacity to draw conclusions upon the results: why have they been obtained? What is the causal link between objectives and results? What should be changed to have better results? Evaluation is not a measurement activity but an interpretative activity.


5. MAB is an advisory body charged under the Public Service Act with providing advice to the Commonwealth Government on public sector management issues. It is chaired by the Secretary of the Department of Prime minister and Cabinet, and its members are drawn from the heads of a number of Commonwealth Departments and agencies, the private sector and the Australian Council of Trade Unions. MIAC advises MAB on technical aspects of public sector management and identifies and promotes best practice.
REFERENCES


Chapter 4

IMPLEMENTATION OF THE
GOVERNMENT PERFORMANCE AND RESULTS ACT OF 1993

by

Walter Groszyk

Office of Management and Budget, United States

The Government Performance and Results Act of 1993 (GPRA) is intended to bring about a fundamental transformation in the way government programs and operations are managed and administered. The United States is at the mid-point of a four-year period for making any needed changes in government processes, systems and practices before the law takes full effect in 1997. This chapter examines the United States experience to date.

Introduction

On August 3, 1993, President Clinton signed into law the Government Performance and Results Act.

The main features of this law are:

♦ A requirement for Federal departments and agencies to prepare strategic plans, beginning with an initial plan to be submitted to the Office of Management and Budget (which is an agency within the Executive Office of the President) and to Congress by September 30, 1997.

♦ A requirement that Federal departments and agencies prepare annual performance plans, setting out specific performance goals for a fiscal year, starting with a performance plan for fiscal year 1999. (The Federal Government’s fiscal year begins October 1 and ends the following September 30. Fiscal year 1999 begins on October 1, 1998).

♦ A requirement that the Office of Management and Budget (OMB) prepare an annual government-wide performance plan, which is based on the agency annual performance plans. The government-wide performance plan is to be a part of the President’s budget and is transmitted to Congress. In the agency and government-wide performance plans, the levels of program performance to be achieved will correspond with the program funding level in the
budget. The first of these plans will be for the fiscal year 1999 budget, which Congress should receive in February, 1998.

- A requirement that Federal departments and agencies submit an annual program performance report to the President and Congress, and which compares actual performance with the goal levels that were set in the annual performance plan. The annual report is due six months after the end of a fiscal year. The first report, covering fiscal year 1999, is to be submitted by March 31, 2000.

- Provisions giving managers greater flexibility in managing by allowing the waiver of various administrative controls and limitations. In return, managers are expected to be more accountable for the performance of their programs and operations.

In the years prior to these government-wide requirements coming into effect, the law provides for several sets of pilot projects. The performance measurement pilot projects test and demonstrate whether the structure and specifications for the annual performance plan and program performance report work as intended. The managerial accountability and flexibility pilot projects examine the practical application of the managerial accountability and flexibility concept on Federal managers and their staff.

Within the Federal Government, GPRA applies to all 14 Cabinet departments, virtually all independent establishments (agencies), and all government corporations. A government corporation is an entity either owned or controlled by the Federal Government. Departments and establishments have component units, which may be called a bureau, administration, or service, and are covered by being a part of a department or independent establishment. In this paper, the term “agency” is hereafter used to mean a department, independent establishment, or government corporation.

The change that GPRA seeks to bring about would place much greater emphasis within the government on what programs are actually accomplishing, and how well the accomplishments match with the programs’ purpose and objectives. In the United States, this is called “Managing for Results”. To this juncture, the primary focus has more usually been on the amount of budget resources obtained, and what these buy: staff on-board, for example, new equipment, supplies, or the number of grants, procurements, and transactions that could be made.

GPRA centres on the role and responsibility of Federal managers (of which there are several tens of thousands within the Federal government). Other reforms have usually focused on particular functions (such as accounting and auditing) and which are often support or ancillary activities in contrast to an agency’s core programs.

Many (perhaps most) Federal officials are mainly evaluated not on what their programs accomplished, but on whether they followed the administrative rules and internal practices that governed their conduct as managers. Under GPRA, this adherence still matters, but one’s role and responsibility for a successful or failing program should be of greater importance.
Part 1: Background

Earlier Efforts

GPRA is not without its historical ancestors. These can be traced over the past 30 years, and, somewhat ominously, largely failed to take root. Indeed, if any had succeeded, this paper would have a different topic and theme.

The first was the Program, Planning, and Budgeting System (PPBS), which was introduced in the 1960s and sought to extend the program management and budgeting scheme being used by the Department of Defence to the government-at-large. What Vietnam-era “whiz kids” in the Pentagon were able to accomplish in the Department of Defence with the operations research focus of PPBS could not be readily replicated elsewhere. PPBS was succeeded in the 1970’s by Management by Objective (MBO) and then by Zero-Base-Budgeting (ZBB). Within the Federal Government, MBO and ZBB soon became artefacts of that decade although an evolved form of PPBS continues to be used in the Department of Defence.

During the 1980s, there were three significant initiatives: productivity improvement, quality management, and a short-lived resurrection of MBO. The productivity improvement effort quickly became mired in difficulty: problems arose in defining what units would be measured, compounded by the widespread incapacity of government financial systems to cumulate and assign costs for the units. Quality management continues to this day, although in an evolving and more focused form.

Will History Repeat?

Is GPRA yet another reform that will flower briefly and soon fade? What makes GPRA different from these earlier initiatives, and why might it succeed when the others have not?

An important distinguishing aspect of GPRA is that it is a law; the others were established through Presidential directive. GPRA is thus a creation not only of the President, but of the Congress as well. Seen another way, laws such as GPRA can be enduring monuments to governance; Presidential directives are more fragile edifices. Indeed, most of the earlier initiatives began and ended with a Presidential term.

By design, GPRA is not trying to transform immediately the whole Executive branch of government. Many of these other efforts attempted to do so, and quickly. The combination of haste and scope became the path of failed ambition. Major change requires time: staff often need to adapt, and systems and processes must be adjusted. Several initiatives sought to expand a much smaller effort that had worked well for a particular type or level of function into government-wide application. A full-scale instant immersion strategy was followed. But in not allowing for lessons learned, experience gained, and a trial and error approach, the onset of problems and flaws in the course of putting the reform in place first bred agency reluctance and then resistance to the reform itself.

In the past, too much was expected from existing systems and processes for measuring, reporting, and analysing information. Only in recent years with widespread use of desktop computers and linked networks of computers, has the collection and presentation of timely performance information moved from being a paper- and staff-intensive effort to something that can be done in a more practical manner.
Performance information that managers use in managing programs and operations forms the underpinning for GPRA performance measurement. Some earlier efforts used this (or similar) information in a more limited way, such as analyses for decision-making. While such uses can be very important, the data was often special purpose, developed and used by only a few staff. These staff had little or no ownership for any goals or commitments associated with the information. Any tie to what managers would do (or accountability for how they managed) was remote, if it existed at all.

Perhaps most significantly, these are different times. Throughout previous decades, funding for Federal programs was generally stable, if not growing. Now, funding is mainly decreasing, and programs are increasingly being challenged to explain their purpose and demonstrate their value. Program performance now counts for much more than in the past.

The public’s view of government has also changed. Several commentators have declared that American government is currently beset with three deficits: a budget deficit, a performance deficit, and a trust deficit. The performance deficit is indicated by a public that doesn’t know what its government is doing. The trust deficit marks a public which has low confidence in its government doing the right thing. More and better information on what the government is doing and how well it is doing can address the trust and performance deficits.

**Immediate Antecedents**

GPRA’s origin can primarily be traced to two separately conceived proposals put forward about five years ago. One proposal was described in the Management Report (January 1989) of then President Reagan. This report contained a chapter, prepared by OMB staff, titled “Government of the Future” and which outlined the basic structure of what would become four years later, GPRA.

The other seed for GPRA sprang from Congress. In 1991, Senator William Roth first introduced the legislation that would eventually become the Government Performance and Results Act. This was the other origin of GPRA. The proposed law was largely framed on the experience of one American city -- Sunnyvale California -- which had successfully been using many of GPRA’s main features for over a decade. The following year, the legislation was extensively rewritten and 19 other Senators (nearly equally representing both political parties) joined Senator Roth in sponsoring the proposed law. In late 1992, the legislation was approved by the Senate, but the House of Representatives took no action. In 1993, newly-elected President Clinton gave early and strong support for the proposed legislation. GPRA was subsequently passed by both the Senate and House of Representatives without objection, and was promoted and supported by members of both political parties.

The Chief Financial Officers Act became law in 1990, and several features of this Act also helped lay the groundwork for GPRA. The first was an instruction to agencies that program performance information was to be included in the annual financial statements required by this Act. (This performance information was retrospective, covering a past fiscal year.) The second was a joint OMB/agency effort which developed common performance measures for over 15 functions similar to those being routinely done in the private sector. These commercial-type functions included areas such as debt and credit management, real property management, electric power generation and distribution, supply and inventory management, and insurance.

The Executive branch did not rely solely on the Sunnyvale experience when it participated with the Congress in redrafting the proposed law, or in assessing whether GPRA would be feasible or
practical at a national government level. Any confidence that GPRA might be carried out successfully primarily arose from two sources: the experience in developing common performance measures for the annual financial statements and the increased inclusion of performance information in these statements; and from similar initiatives that were taking hold in other countries. From a U.S. perspective, much weight was placed on a conclusion that these other national initiatives were still being advanced a half decade or more after being started (most of the previously referenced Federal efforts had already languished within such a time-period) and overall seemed to be succeeding. The OECD Public Management Committee was especially useful in identifying these other national initiatives.

Nothing in GPRA is new or revolutionary to the American experience. GPRA’s major elements can be found in most businesses, where these have been used for decades but for a different ultimate objective: profit or loss. What GPRA seeks to create is the government’s counterpart to the business “bottom line”.

**Parallel Efforts**

As GPRA became law, other initiatives were also coming to the forefront of Federal governmental reforms. The most prominent of these is the National Performance Review (NPR), a wide-ranging look at how government could be made to work better, and led by the Vice President.

The NPR has embraced GPRA’s purpose and concept. Several parts of the NPR relate directly to performance. These include:

- **Customer service standards.** Agencies have established about 2000 standards. The customer service standards reflect a major Federal emphasis on improving the delivery of government services and products to their users and beneficiaries. The customer service standards continue an emphasis on quality improvement and management that began last decade. Not all the standards are measurable. In time, those that are should be incorporated as performance goals in GPRA plans and reports.

- **Performance agreements.** These are agreements between the President and Cabinet Secretary or the head of an independent agency. The concept is based on performance-based employment contracts that are being used in several other countries, with several important differences. These Presidential agreements are at the highest level, and not between a minister and an agency director, which is the usual starting point elsewhere. The agreements are also not used to reward or penalise individual performance. Based on the Presidential agreement, agencies are then to create other agreements between the agency head and subordinate officials, and which “cascade” down through the organisation. About ten performance agreements have been signed to date. Some of these have cascaded down through an agency, and at these tiers, can be linked to personnel evaluation systems and pay incentives. The intention is to expand the number of agreements, and include more performance-related goals and objectives that are in an agency’s strategic plan or annual performance plan.

- **Reinvention labs.** Over 200 reinvention labs were created in government agencies. These labs are testing ways to streamline administrative processes, and reduce and eliminate
unnecessary controls on Federal managers. The main focus of the labs has been on controls and rules established by the agency itself, and not on government-wide requirements.

♦ **Performance partnerships.** Performance partnerships are negotiated agreements between the Federal Government and individual States or local governments, in which the Federal Government provides the State or local government with greater flexibility in administering a program in exchange for greater accountability for program performance and results. Apart from social security and veterans programs, the Federal Government directly delivers few benefits or services to the public. More often, the States or local government provide these services, with the cost being funded entirely or partly by the Federal Government. Performance partnerships can bring more attention on the value and impact of these Federal funds. At this time, a small number of partnerships are being negotiated.

Some assert that there is no more important measure to be included in an annual performance plan than the cost per unit of service, unit of output, or unit of result. However, existing financial systems in many government agencies lack the capacity to assign, cumulate, or report costs. The government is addressing this shortcoming in two ways. First, the Federal Accounting Standards Advisory Board has recommended and OMB has issued an accounting standard requiring cost accounting in the agencies. Second, the government is completing a definition of the financial system capabilities and improvements that will be needed to achieve this cost accounting standard.

**Part 2: The Main Elements of GPRA and their Status**

An overview of the design of GPRA reveals several underlying premises and aspects:

♦ Virtually the entire Executive branch is subject to the Act. Provision is made for classified plans and reports from several agencies where the information within must be protected for national security reasons.

♦ Fully meet GPRA requirements number about 75. Approximately 60 smaller agencies (those with $20 million or less in annual spending) may be relieved by OMB from having to meet all the requirements.

♦ GPRA is a very large umbrella, covering not only the expenditure of moneys for programs and operations, but regulation and tax expenditures when the latter relate to the achievement of performance objectives. (Tax expenditures are allowances in the tax code, such as credits and deductions, and represent revenues foregone.)

♦ Because there is neither a single nor uniform approach to performance measurement across the government, GPRA avoids over-specification or rigidity in how agencies meet the Act’s requirements. Flexibility is key if agencies are to adapt and fit their own particular processes and activities into a larger performance framework.

♦ Almost everything that is done by the government can be measured in some way. Certainly, there are a few activities which are not worth measuring from a performance standpoint (interest payments on debt, for example), or where an overly finite level of measurement makes little sense. And everything will never be measured equally well. To avoid unproductive debate on how best to categorise or classify any particular measure, GPRA
defines only two types of performance measures: output and outcome measures. Both were defined sufficiently broadly to cover other measures, such as impact, attribute, efficiency, etc.

♦ The Act does not establish or assign responsibility for GPRA implementation to a particular organisation or unit within an agency. In the law, the individuals given responsibility for carrying out GPRA are the Director of OMB and the agency heads.

♦ Agencies should not create a new GPRA bureaucracy. Agencies are admonished not to “reinvent the wheel”, but to use and apply, to the extent appropriate, existing processes and products when developing the plans and reports required by GPRA.

♦ Plans and reports are to be brief and concise. Hectares of forest should not be cut to prepare voluminous piles of paper, and which only intimidate the prospective reader.

♦ Federal agencies do not generally require a substantial investment in new systems or procedures to manage programs effectively. To agencies seeking investment funding for this purpose, two questions should be raised: have managers regularly examined how their programs were doing, and acted on this information? and, what measures are they using for this purpose? Few agencies are likely to concede that either their managers have never managed, or used a fundamentally flawed set of measures in managing. While there are legitimate needs to improve and expand the set of measures, or to collect better data more quickly, these should be a small fraction of current agency budgets.

♦ The Federal budget traditionally includes descriptive material showing the relationship between spending levels and program activity or achievement levels. However, this is not performance budgeting in the GPRA context. GPRA requires a third set of pilot projects (during fiscal years 1998 and 1999) which test performance budgeting. In these pilot projects, optimisation analyses will be done, presenting choices and trade-offs between different levels of performance for the same or different budget levels. Cost accounting data will be needed. Because of substantial uncertainty about how well the performance budgeting pilot projects will work, or interest in having budgets crafted in such a fashion, GPRA defers until the next century a decision on whether the government should produce a budget based on these performance budgeting principles.

**Strategic Plans**

GPRA requires that an agency’s strategic plan contain the following elements:

♦ a comprehensive mission statement;

♦ a description of general goals and objectives and how these will be achieved;

♦ a description of the relationship between performance goals in the annual performance plan and the general goals and objectives in the strategic plan;

♦ an identification of key external factors that could affect achievement of the general goals and objectives; and

♦ a description of program evaluations used, and a schedule for future evaluations.
Strategic plans cover a minimum of six years, and are to be revised and updated at least every three years.

Strategic plans provide the foundation for carrying out all other GPRA requirements. The strategic plans:

- state an agency purpose: why an agency and its programs exist, and what will be accomplished and when.

- define the long-range course of the agency with sufficient precision to guide the short-term actions of agency managers.

GPRA does not specify that a government-wide strategic plan be prepared, nor are multi-agency plans for cross-cutting functions envisioned. The absence of such plans dictates that appropriate co-ordination be done among those agencies whose activities and operations form part of a larger inter-agency program. In these instances, the objective is to assure that individual agency goals are synchronised and harmonised with related goals in the plans of other agencies. The agencies are working jointly to develop a common, consistent approach in setting goals for the inter-agency programs. Joint groups already started cover areas such as trade, research and development, land and resources management, and credit programs.

Development of strategic plans should be used as the time for broad discussion about the agency’s future programmatic direction and priorities. Consultation with Congress is required during plan development, and agencies are also to seek the views of other interested or potentially affected parties. By minimising, if not avoiding, a recurring annual debate about program direction and purpose, program managers are given a stable basis from which to set specific annual goals in the performance plans. Strategic plans should not be a chosen means for raising basic policy questions, such as altering eligibility requirements for beneficiaries.

While full agreement or general consensus on the strategic plan content is to be hoped for, this may not always happen. Agencies may have to make difficult choices from among competing or differing views. To underscore both the importance of consultation and to record that all relevant and germane views were considered, agencies are to summarise the consultation that occurred and any contrary views received during plan preparation.

Some agencies have years of experience in preparing strategic plans. Others have yet to write their first. But no previously prepared strategic plan has yet been identified as meeting all GPRA requirements. While several existing plans already appear to meet some of the required GPRA elements, such as a mission statement and a well-defined set of general goals and objectives, what generally is lacking is coverage of external factors, the relationship between the strategic plan and the annual performance plan, and satisfying the required consultation with Congress. The experienced agencies are likely to re-examine their role and missions in a fundamental way when developing GPRA strategic plans, and not simply re-package existing plans.

Annual Performance Plans

An annual performance plan contains the following elements:
♦ One or more performance goals for each of the programs and operations covered in the plan. (A performance goal is a target level of performance expressed as a tangible, measurable objective against which actual achievement can be compared, including a goal expressed as a quantitative standard, value, or rate. Also, a plan need cover only the major programs and operations in an agency).

♦ Performance indicators that will be used in measuring outputs and outcomes. (A performance indicator is a particular value or characteristic, and is used when a goal is not self-measuring).

♦ A description of the means to be used to verify and validate measured values.

♦ A brief description of the operational processes, skills, and technology, and the human, capital, information, or other resources required to meet the performance goals.

♦ Any proposed waivers of administrative requirements and controls to give managers greater flexibility.

In choosing the goals and indicators to be included in the plan, agencies should be guided by the following principles:

♦ Goals and indicators should primarily be those used by program managers to determine how well a program or activity is doing in achieving its intended objectives.

♦ Include measures that will be useful to agency heads and other stakeholders in framing an assessment of what the program or activity is accomplishing.

Performance goals in the annual performance plan are derived from the general goals in the strategic plan. As such, a performance plan defines the incremental progress in a particular year toward achieving or sustaining the strategic plan’s goals. Often, the goals in a performance plan will be outputs, while the general goals are more likely to be outcomes.

Many of the general goals in a strategic plan will not be fully achieved until years in the future. Quite probably, because of shifting tenure and turnover, many of those setting these goals will not be in positions of direct responsibility when achievement ultimately occurs. The 12 month timespan of the annual performance plan allows the government to fix accountability for one year’s progress and achievement on individuals and organisations with the reasonable expectation that managers will hold their position over this period.

Performance measurement pilot projects are being used to demonstrate whether agencies can prepare annual performance plans that meet GPRA requirements, and to identify any difficulties agencies are encountering in plan preparation. (There are no pilot projects for GPRA strategic plans).

Pilot projects were officially designated in all 14 Cabinet departments and an equivalent number of independent agencies. Some agencies have more than one pilot project, and over 70 individual pilot projects are covered by these designations. In some instances, agencies have supplemented the designated list with their own informal or internal pilot projects.

The pilot projects range from the very large -- the entirety of the Social Security Administration and the Internal Revenue Service -- to the very small. As pilot projects are a learning experience for the agencies, few restrictions or limits were put on who could be designated. Nearly 450 000 civilian
and military employees are covered by these pilot projects. While this is a large scale endeavour, a more critical facet of the designations is that the pilot projects cover all but two of the 30 major functions of government. The two functions not covered are direct delivery of medical and nursing care, and electric power generation and distribution.

Depending on when they were designated, the pilot projects have submitted up to three separate annual performance plans. With the final set of plans submitted in the Spring of 1995, assessments are now being done on both the agency experience and the overall quality of these plans. Some initial impressions:

♦ Enough pilot projects were able to produce performance plans of sufficient quality and covering a variety of functions to prove, that good plans can be produced across the government. Expectations are that a fourth or more of the individual pilot projects will be selected as illustrative examples of promising practices. The examples are not perfect plans, but a selection of parts from the better plans which other agencies might consider adapting or emulating in their own plans.

♦ A significant number of pilot projects encountered substantial problems. About ten percent of the pilot projects were eventually withdrawn. Reasons for withdrawal included inherent difficulty in defining program goals, reorganisation of the program or staff, and the failure of other parties (outside the agency) to produce the performance information that was needed to prepare the plans and the program performance reports.

♦ Many of the plans contained weak performance measures. These measures included statements purporting to be goals, but which were no more than inherently unmeasurable declarations; goals where performance could not be calculated or determined because one or more values were omitted (much like a formula with only one integer); goals which described the procedural steps being taken to set future goals and measure performance; and goals that were unrelated to the main purposes of a program.

A concern that the poorer plans might be indicative of future plan quality when plans are submitted “for real” in 1997 has prompted OMB to take certain steps. These actions include:

♦ An orientation for all OMB professional staff on performance measurement, to better prepare these staff in assisting agencies in selecting appropriate and useful performance measures.

♦ An accelerated effort to have agencies develop and provide more and better performance information with the fiscal year 1997 and 1998 budgets. In the Spring of 1995, OMB conducted a Spring Review on performance in which the availability and value of performance information for some key government programs was assessed. Giving consideration to having the fiscal year 1998 budget process be a “dry run” for fiscal year 1999, when GPRA requirements take effect. Agencies have been alerted to expect early calls for providing some parts of the strategic plans in the Summer of 1996, and full descriptions of the performance measures agencies expect to use in their fiscal year 1999 performance plans may be submitted with the fiscal year 1998 budget. This timetable gives OMB and the agencies at least a year to forge agreement on the measures that will be used in fiscal year 1999.
Renewing the emphasis on agencies quickly beginning to collect current performance data so that baseline and trend information is available when the specific performance targets for fiscal year 1999 are set. Without baseline information, calculating the planned performance levels can be little more than guesswork.

**Annual Program Performance Reports**

The annual program performance report should contain the following elements:

- A comparison of the actual performance achieved with the performance level(s) specified for each performance goal and performance indicator in the annual performance plan.

- If a performance goal was not met, an explanation of why the goal was not met, along with either:
  - the plans and schedules for achieving the performance goal in the future; or
  - a statement that the performance goal established is impractical or infeasible, and expressing the agency’s intention to modify or discontinue the goal.

- The summary findings of any program evaluations completed during the fiscal year.

- A description of the use, and an assessment of the effectiveness, of any waiver of administrative requirements and controls in achieving performance goals. (This is the managerial accountability and flexibility portion of the report).

By completing one cycle of planning, doing, measuring, and evaluating, and helping begin a new cycle, program performance reports close the loop. What programs (and managers) achieved or failed to achieve is factored into the next round of plans, both performance plans and strategic plans. The adjustments, revisions, and corrections that flow from information in an annual report renew the plans, making them current and real. The report’s retrospective scrutiny of what was actually achieved also serves as a self-correcting device. Goals found to be too ambitious will be lowered, and goals which consistently underestimate what really can be accomplished will be raised.

As the reports are sent to the President and Congress, and are available to public, accountability and performance will become quite visible -- perhaps uncomfortably so.

The performance measurement pilot projects are also being used to demonstrate whether agencies can prepare annual program performance reports that meet GPRA requirements, and to identify any difficulties agencies are encountering in preparing these reports.

Pilot projects that prepared a performance plan for fiscal year 1994 have submitted program performance report for that year. These reports are currently being reviewed. In most instances, the quality of the report reflects the quality of the plan. In some instances, agencies are able to report actual performance even though performance goals in the performance plan lacked a measurable value.
Managerial Accountability and Flexibility

The conceptual entirety of GPRA is fused from three basic elements: (1) strategic planning, (2) performance measurement (which covers goal-setting, measurement, and reporting), and, (3) greater flexibility for managers in return for greater accountability for results.

This third element -- managerial flexibility -- is introduced by waiving selected administrative requirements and controls. These administrative and procedural requirements are prescribed by agency rules and directives. (GPRA gives no new authority to waive administrative requirements established in law.) Requirements and controls that may be waived are those imposed across the government by an agency which has a central management role, and cover areas such as financial management, personnel, supply, buildings, etc. The Department of the Treasury, the General Services Administration, OMB, and the Office of Personnel Management are considered to be central management agencies.

A key conclusion reached after examining the experiences of other governments with reforms similar to GPRA is that incentives are essential if managers are to change how they manage and become more accountable. Fewer rules, fewer controls, and more latitude in how one uses and directs resources and organisations to accomplish program and policy goals were seen as key incentives. Recognition of superior performance through individual financial reward is another. (Imposing a financial penalty for poor performance is more difficult under the current system.)

While the managerial flexibility element of GPRA attempts to mirror what other countries are doing, a manager’s autonomy under GPRA is likely to be less than the discretion given his or her counterparts in some other countries, where they can decide how and on what to spend on running costs, or to set compensation rates.

A set of managerial accountability and flexibility pilot projects are to test how important this third element could be in transforming how agencies are managed and programs are administered. No pilot projects have yet been chosen, but review of eight nominated pilot projects and their requested waivers will be completed in October 1995.

On the whole, the eight nominations raise uncertainty about whether the pilot projects will fully test the managerial accountability and flexibility concept.

The nominations are a mixed lot. Several nominations are very narrow, seeking only to waive a single requirement, some are a mixture of potentially significant changes and requirements which are more an annoyance than a substantive burden. Some prospective waivers have been overtaken by time (the requirements to be waived no longer exist, see the reference to the National Performance Review below). Other requested waivers are tangential, at best, to the notion of managerial flexibility.

While it is too early to draw any conclusions, some speculative explanations have been voiced about why this set of pilot projects is turning out this way. These include:

✦ The major downsizing of the Federal workforce. Still underway, this downsizing may have made agencies uncertain about whether relief from limits on how many staff an agency could employ was appropriate or needed.
A pre-emptive de-regulation by central management agencies. GPRA was enacted prior to completing the first phase of the National Performance Review. Several agencies, most particularly OPM, have cut substantially the requirements imposed on other agencies. One can’t waive what is no longer required.

Agencies may have discovered that the most burdensome requirements and controls are self-inflicted, and not imposed by central management agencies.

Perhaps, the Federal Government may be an exception to the prevailing principle that managers need incentives, such as flexibility, if they are to be more accountable for results. Or perhaps, it is still too early for most managers to determine what they would like in terms of added flexibility.

Some agencies may have found that developing good performance measurement pilots was task enough, and the managerial flexibility pilot projects were overloading their capacity at this point in time.

Part 3: General Observations

Some General Findings

The concept and basic structure of GPRA appears to be sound. No fundamental flaws have yet been discovered. This should not be surprising, because GPRA concepts are neither new nor unproven.

Enough agencies are making enough progress to indicate that GPRA can be done. However, GPRA is still in a formative stage, and the reality of what this law actually means for agencies, managers, the Congress, and the public is yet undetermined.

A few agencies are beginning to introduce new standards for evaluating the performance of senior managers and officials. These new standards place much greater weight on how well managers are managing their programs, and on what programs achieve. Changes in how managers are evaluated and rewarded is critical to affecting any change in how they manage.

The large-scale cutting of red tape and unnecessary rules by the National Performance Review is forming a basic underpinning for managerial accountability and flexibility. Personnel and procurement are two areas where major reductions are occurring.

Scaling up to government-wide implementation is not that great a mountain to climb.

About 25 percent of the Executive branch employees are covered by the performance measurement pilot projects when Postal Service personnel and members of the armed forces are removed from the calculation.
Some General Expectations

GPRA will be successful by disappearing! The transformation set in course by GPRA will be completed when agencies and managers carry out its principles and practices as every-day routine, and not because a law requires them to do so.

Implementation of GPRA should and will be uneven across the agencies. If GPRA is easily or perfectly carried out everywhere, then this Act has brought no great change. As a rising tide lifts all boats, time, experience, and comparison should raise the quality of all agency plans and reports to a satisfactory level.

Perfection is the enemy of the good. Agencies should strive to do the best they can now. To delay until an elegant design or process is put in place, or until all unknowns are explained, simply means nothing will ever happen. Public and Congressional expectations about the value and usefulness of plans and reports are likely to be modest at first, and grow as better examples appear. Substantive plans and reports, even if incomplete, will likely fare better in these forums than those which are seen as without substance.

There is no risk without failure. Managers and agencies will make mistakes. Not all strategies will work; every program will not succeed. GPRA blends accountability with entrepreneurship. Without some tolerance for errors, managers will be risk-avoiders. A natural tendency to focus on faults and failings must be capped. Some mistakes--not all--should be forgiven. Some may lead to changes in managers. A few mistakes will lead to corrective steps within the systems and processes. If the answer to every failure is to place more controls and bounds within the management structure, then the managers remaining in government will be those who view their sole responsibility to be that of following rules.

Collectively, managers must have a sense of stewardship for their programs. (In other countries, this proprietary interest may be termed ownership). GPRA rides on the shoulders of many thousands of Federal managers, on their conclusion that what GPRA seeks to do is a help, not a hindrance. GPRA brings fundamental change, and change is not always welcome. Some managers will be uncomfortable, others may resist, but initial indications are that many managers view GPRA as a challenge and opportunity. GPRA probably asks more from them, but also allows them to demonstrate their managerial skill and capacity, and to highlight the value and worth of the programs and operations they oversee and direct.

Agencies and OMB must link GPRA plans and reports to budget formulation and execution processes in a decisive way. OMB's recently issued instructions on the preparation and submission of strategic plans underscore this principle.

Some General Concerns

An attempt to make GPRA a budgetary axe would confirm the worst fears of agencies and managers that this Act is not about changing how programs are managed, but about fewer programs and less money. Certainly, as program performance is reviewed, programs that achieve little or nothing may be eliminated. But not always. With changes, some failing programs can succeed. With more funding, underperforming programs may do much better. Managers should believe that the evaluative judgements arising from GPRA will produce remedies as well as terminations.
Managers are frequently reluctant to be accountable for what they do not control. In many Federal programs, the States, local governments, or other parties administer the programs and provide the services. (This year, Congress may lessen further the requirements and standards that Federal agencies can establish for some of these programs). Actions of these other entities can lie beyond the span of control of a Federal manager. If not within a manager’s span of control, such programs are usually within his or her span of influence. The general goals in a strategic plan will cover programs and activities within both span of control and span of influence. The bridge not yet done is how to bring performance for span of influence activities into the annual performance plans. The performance partnerships mentioned earlier can be one means, and several of the performance measurement pilot projects are joint Federal-State efforts. In these pilot projects, the States are setting goals for their own performance. Collaboration and co-operation will be key. A complicating concern is a perception that this bridge might create a future national, performance-based management system of gigantic proportions, involving and affecting Federal agencies, States, local governments, and others. There are no plans for such a system; indeed, it would be resisted as infeasible if not impractical. More critically, the idea that Federal managers would, in some manner, set performance levels, such as service standards, for State and local governments is untenable politically. The prevailing view in that Federal mandates should decrease, and mandates that are not Federally funded should disappear.

A lack of Congressional interest is likely to be fatal. What is not used eventually is no longer produced. ‘If and how Congress will use GPRA plans and reports is a test to come.

Past, current, or future program performance must become a factor in funding choices and decisions. If the government or public view is that simply spending money on a program or problem is performance enough, than the effectiveness or efficiency of programs (their real performance) is of little consequence.

Since GPRA became a law, other performance-related initiatives have got underway. Others began earlier or loom on the horizon. While this might become too much of a good thing in this still-formative stage of GPRA implementation, the greater peril is that many separate, parallel efforts could fragment organisational responsibilities and staff, duplicate products, waste effort and time, and confuse. As Benjamin Franklin, America’s emissary to France during the War for American Independence, stated, “We must all hang together, or assuredly we shall all hang separately.” OMB is developing an overall framework that would consolidate various plans and reports and link these with processes for forming and executing the budget. This framework could possibly encompass a dozen or more of the current performance-related efforts, including GPRA and other initiatives (such as financial management and those of the National Performance Review previously referenced).
Chapter 5

IMPLEMENTATION OF GOVERNMENT PERFORMANCE AND RESULTS ACT 1993

Commentary by

Richard Boyle

Institute of Public Administration, Ireland

Introduction

The GPRA represents an interesting and comprehensive approach to the promotion and development of performance management in government departments and agencies. I would like to focus in my remarks on making some comments on the three ‘legs’ of the stool on which GPRA is based: strategic plans, annual performance plans and reports, and managerial accountability and flexibility.

Strategic Plans

The GPRA requires Federal departments and agencies to prepare strategic plans. This is in line with a recent emphasis on strategic planning and management in OECD government initiatives e.g. the Strategic Management Initiative in Ireland, the requirement of Next Steps executive agencies in the UK to produce Framework Documents, the use of corporate plans in the Australian Public Service. Such initiatives are a welcome recognition of the need to develop and enhance strategic management capabilities in government departments.

There are indications in the paper that the GPRA stresses strategic planning rather than strategic management: the requirement under the Act is the production of a strategic plan. The danger of this approach is that form may come to dominate substance. Whilst a plan may be needed — to provide a framework for actions — in many ways it is the process used to arrive at the plan that is crucial.

Strategic planning tends to focus on goals and objectives, with an emphasis on the linear extrapolation of trends. In this rational planning model, strategy is often seen as the responsibility of senior management in the organisation, aided by a strategic planning unit or the like. Strategy is then cascaded down the organisation. However, this can lock people into the pursuit of a grand design. In reality, strategy is often evolutionary, with an emphasis on developing strategic thinking rather than strategic planning being the key to success (Jackson, 1995).
In this alternative approach, strategy development is a continuous process, not a once-off event that occurs at particular time periods. Plans are the by-product of strategic thinking in government departments and agencies rather than the end product. If the planning system become too rigid, with only periodic “formal” reviews, then inappropriate and out-of-date strategies can be the result.

In practice, if the strategic plans advocated under the GPRA are to avoid this danger, it will be crucial that lessons are learnt not just about whether plans meet GPRA requirements (such as having a mission statement, setting goals and objectives, identifying key external factors and outlining programme evaluations) but also about how the plans are produced and used. For example, how are staff and clients involved in the planning process, what role do line managers and support staff play etc.? Good practice lessons must focus on the process of establishing plans as much as on promoting the improvement of plan content.

A particular issue worthy of attention will be the structures established to meet the strategic planning requirements. If strategic thinking is the guide, aimed at informing an organisation’s daily actions, then strategic management becomes the job primarily of top managers and line managers. But this puts particular pressures on already hard-pressed managers. There is still a need for strategic planning staff, but as facilitators and integrators rather than “doers”. Their role should be to help managers work through the strategic management process, providing support and guidance.

A notable feature of the strategic plans proposed under the GPRA is that they should contain a description of programme evaluations used, and a schedule for future evaluations. This is a welcome and innovative feature, attempting to close the planning — implementation — feedback policy loop. It will be interesting to see how this element works in practice. One concern is that Federal-level evaluation in particular has come in for scrutiny and criticism in recent years. Evaluation has been seen in some circles as one of the “control” instruments imposed on front-line administrators and stopping them doing a good job (Caroll, 1995). In this scenario, freeing managers to manage means fewer evaluations. Yet fewer evaluations mean less information available at the strategic planning stage when determining where to next, in the context of what has and has not worked in the past.

**Annual Performance Plans and Reports**

One particularly striking feature about the required performance plans and reports is the systematic piloting of experience prior to “going live” with 1999 plans and year 2000 reports. Piloting of major initiatives in an extensive and systematic manner is a rare feature in public administration. It would be useful to know more about the piloting experience: how pilot sites were chosen; how the lessons learnt are to be assimilated and disseminated, and so on. In general, learning from the strengths and weaknesses of this piloting approach could improve understanding of this approach to managing and introducing change.

The paper states that goals and indicators used in plans and reports will be those used by programme managers to determine how well a programme or activity is doing. It also states that the goals in a plan are more likely to focus on outputs than outcomes. A strength of this approach is that the goals and indicators should be seen as directly of interest and use to managers, and not an imposed burden which they must complete before getting back to their real work.

A possible danger, however, is that the client or citizen perspective is lost or downplayed. Managers’ goals for a programme or activity may or may not coincide with those for whom the programme or activity is being provided. Similarly, the focus on outputs may mean that whilst
judgements on efficiency can be made by studying plans and reports, judgements on effectiveness (outcomes) may not prove so easy.

These concerns are likely to be particularly crucial in how the plans and reports are seen and used by Congress. The paper notes that a lack of Congressional interest is likely to be fatal to the GPRA. Their interests are more likely to be with effectiveness and with the users’ view of service provision. If these issues are not adequately addressed in the plans and reports, then Congress may lose interest in them.

The requirement of reports to include information on programme evaluations offers one mechanism for ensuring that such concerns are not ignored in the plans and reports. Evaluations should give information on outcomes and the users’ perspective and enable this data to be built into the process. It is notable that the Australian Public Service experience with performance reports is that where evaluation has been carried out, performance reports are seen as being of better quality (Task Force on Management Improvement, 1992). Evaluations require clarification of objectives, definition of measures and indicators, and the development of relevant information systems.

However, getting Congressional attention and interest in performance reports and plans is not on its own a guarantee that everything will work out smoothly. Congressional interest in goals and indicators is likely to be for accountability purposes, with indicators being used as an aid to holding organisations accountable by those for whom the work is being done. But the most common performance indicators in government departments are descriptive indicators; they do not on their own provide answers about what is good or bad performance, rather they raise certain questions. Descriptive indicators are more suitable for development purposes i.e. as a means of learning for managers and staff, than for accountability purposes (Boyle, 1996). It is likely that Congressional users will have to be assisted in interpreting and using such indicators, for example, by using a set of indicators acting together rather than single indicators as a sign of good or bad performance.

Managerial Accountability and Flexibility

Managerial flexibility – the third “leg” of the stool of strategic management, performance measurement and managerial flexibility – appears from the paper to be the weakest leg at present. The paper notes that nominations for pilot projects in this area has been mixed, without a clear focus. It will be important to determine which of the potential reasons given for this outcome in the paper are the real reasons. If departments and agencies feel they already have enough flexibility, this requires quite a different approach than if they are feeling overburdened with initiatives or if flexibility is not deemed relevant when downsizing.

Another incentive, besides flexibility, is the use of good practice and “peer-pressure” to encourage the widespread take-up of good practice. OMB will have a key role to play here. They will be able, as with the pilot projects, to identify and disseminate examples of good practice, and encourage others to emulate these standards. An extensive use of seminars, papers, workshops and the like will be needed to communicate and encourage good practice.

More generally, whilst the goal of improved managerial flexibility is clearly worthwhile and needed to win support for a greater performance orientation, its operation in practice may be questioned by managers in departments and agencies. GPRA could be interpreted as increased centralisation and a growth in the control function of OMB. Some managers may see it as more red
tape and bureaucratic instructions, increasing Congressional oversight powers. Again, how the system is run, the process behind the initiative, will be as crucial as the actual outputs and outcomes.

Conclusions

I conclude then with some final comments and questions. The GPRA as outlined in the paper represents a thorough and thoughtful approach to performance management in government. Particularly notable features, and ones worthy of further discussion are:

a) The fact that legislation is the institutional anchor for the initiative. This obviously gives standing to the initiative and ensures it will be taken serious. It could, however, also impose rigidities and conformance by requirement rather than by willingness.

b) The use of pilot projects and long-time frame for development recognises that such initiatives take time and that there is a need to systematically learn from experience. How to maintain interest and stimulus over such a time period, and how lessons learnt are disseminated will be key issues to be tackled.

c) The mix of a “top-down” approach to providing the broad framework with a more “bottom-up” approach to the actual production and design of plans, indicators and reports. This should encourage ownership at the local level by managers. The challenge will be matching this approach with Congressional and citizen needs and interests.

d) The key role for OMB in overseeing and driving the initiative. Linking the process with budget decisions and authority clearly enhances departmental and agency interest in and commitment to the initiative. The challenge here will be to ensure that there is an acceptable balance reached between the control functions of a central agency and the development functions of spreading lessons learnt and promoting good practice.

In all, the GPRA is a comprehensive and attractive approach to performance management. It aims to practice what it preaches, by learning from experience both at home and elsewhere, and attempting to put the learning from these lessons into practice. Its implementation in practice will be a major challenge. But there is a good base from which to start.
REFERENCES


Introduction

Measurement of the performance of government organisations is not a completely new phenomenon. Over the past two decades many Western countries have experimented with various systems for linking budgets with performance and with several forms of policy evaluation. In many cases, however, these approaches — like the Planning Programming Budgeting System (PPBS) in the United States and Programme Analysis and Review (PAR) in the United Kingdom — did not achieve the hoped-for results because of deficiencies in implementation strategies (OECD, 1994).

Another important factor in successful use of performance measurement is clarity of the conceptual framework. With respect to this criterion Osborne and Gaebler mention three basic lessons that government organisations in the United States have learned in developing their performance measurement systems. These are the vast difference between measuring output and outcome, the vast difference between measuring efficiency and effectiveness, and the important difference between programme outcome and policy outcome (Osborne and Gaebler, 1993).

In the case of performance measurement in the central government departments in the Netherlands both implementation and conceptual aspects appear to be of equal importance for success. In this paper recent experiences in this field are described. The focus is on performance measurement in the form of systems of statistical measures that provide insight into Government’s policy performance. Performance measurement in the form of ex ante and/or ex post evaluation of effectiveness of specific policy programmes is only marginally discussed.

We start with a sketch of the early developments in performance measurement in the context of performance budgeting in the Netherlands. In the following paragraphs a number of important instruments to improve performance measurement are discussed. Thus methodology, pilot projects of the Ministry of Finance and the project of the Ministry of Home Affairs on key indicators of public administration, interdepartmental co-ordination and training are reviewed. In addition special issues
such as political support and performance measurement by agencies are discussed. Finally some recommendations are made.

The Case of the Netherlands (1976-1995)

Performance budgeting was one of the early manifestations of performance measurement in the central government of the Netherlands. It was introduced into the central government departments through the Government Accounts Act of 1976 (Sorber and Schild, 1988).

The main functions of performance budgeting were:

- to increase Parliament’s insight into the budget estimates to improve decision making on the appropriation of funds (the allocation function); and

- to foster efficiency of the various policy programs (the management and control function).

Since 1976 the Ministry of Finance’s emphasis on developing and implementing performance budgeting in the central government departments has caused a steady increase in the number of budget proposals explained in terms of performance measures and/or indicators’. However, in the early years, the performance information was largely restricted to data on inputs.

In 1988 the General Accounting Office conducted a study on the quantitative and qualitative aspects of performance budgeting. The general conclusion was rather disappointing. Partially due to the lack of interest of Parliament, ministers and top management, both the quantity and the quality of performance measures published in the budgets were rather poor. This report initiated some further changes by the Government.

The fourth amendment to the Government Accounts Act, which came into effect on January 1, 1992, has further emphasised the relevance of performance budgeting. It required departments to provide in the explanatory memorandum:

- relevant information on factors underlying the budget estimates; and

- if possible, information on the (intended) volume of activities, outputs, and effects of the various policy programs.

The new approach was introduced in the budget of 1989, three years before the amendment officially came into effect.

In the Spring of 1992, the Ministry of Finance conducted a study on the state of the art of performance measures as presented in the explanatory memorandum in the 1991 and the 1992 budgets of the ministries. This study involved an analysis of the information provided on inputs, activities, outputs and effects of each ministry’s budget items. It was concluded that on the average 50 per cent of the budget expenditures was explained, in one way or another by performance data. However there were large differences between the various departments. Explanatory percentages ranged from about 80 per cent to about 4 per cent. Moreover most information still had to do with providing insight into budget estimates and very little information was presented on efficiency and/or effectiveness aspects.
The general conclusion of the evaluation study was that the requirements of the fourth amendment to the Government Accounts Act were not yet fully realised and that more effort was needed in their pursuit.

Thus the Ministry of Finance launched a new initiative for the development of measures of which the main characteristics are that rules were set for the presentation of the measures in the explanatory memorandum of the budgets, that a guideline was issued specifying the measures and that the onus of proof of the impossibility of developing certain measures was laid with the departments.

**Development of Methodology**

As stated by Osbome and Gaebler an important necessary factor for successful use of performance measurement is the clarity of the conceptual framework. In the case of performance measurement in the central government departments in the Netherlands due attention was paid to these conceptual aspects (Ministry of Finance, 1994).

In developing systems of performance measurement three different goals can be identified:

a) providing insight into the budget estimates;

b) providing insight into efficiency aspects;

c) providing insight into effectiveness aspects.

With respect to information on the **budget estimates** the most relevant questions are:

♦ What are the relevant price and quantity (volume) components of a budget item.

♦ Where a budget item covers more than one organisational unit or programme into what more or less homogeneous price and quantity-componentsshould it be split up?

With respect to **efficiency** the following questions are essential:

♦ What is being measured? This question is subdivided into:

  0 What is the relevant level of aggregation of the production process?

  0 What is the primary production process?

  ◊ What (combinations of) elements — inputs, activities and outputs — will be measured?

♦ How is the measurement done? Which methods are used?

♦ How is the measurement of qualitative aspects done?

♦ How are the measures judged? Which standards are used?

As far as **effectiveness** is concerned, the key questions are:

♦ What is the total effectiveness, i.e., total results in terms of the programme objectives?
What can be said about the net-effectiveness of the programme i.e. the extent to which results are due to the programme?

Thus the general purpose of performance measurement is to gain insight into a number of relevant aspects of a production process, a policy process or a budget process to be able to “control” these processes. In this respect four functions of performance measurement can be distinguished:

- providing early warning on developments, especially output and outcome;
- improving the allocation of resources;
- improving the efficiency and effectiveness of production and policy processes: and
- improving accountability, especially in the case of contract management and agencies.

The general approach to performance measurement in terms of efficiency and effectiveness assumes that each production and policy process can be subdivided into four basic elements: input, throughput, output, and outcome. Outcomes (effects) are not part of the production process; they refer to the impact on society. These elements should be measured with the help of performance measures. A performance measure can be defined as a quantified description of a component of a production process, a policy process or a budget process.

Which measures are preferred is determined by the information needs of the decision makers. To prevent misunderstandings clear definitions should be used. In evaluating performance, measures are used for comparisons over time, comparisons with other organisations (cross-sectional analysis) or comparison of actual performance against standards. In this way relevant information on the quantitative and qualitative aspects of the process under consideration can be provided. The approach used is determined to a large extent by the availability of data (Ministry of Finance, Platform Policy Analysis, 1990).

**Implementation (of improvement) of Performance Measurement**

**a) Pilot Projects**

*Initiatives by the Ministry of Finance*

An interdepartmental pilot project on performance measurements was started by the Ministry of Finance in 1990 (Kuipers and Sorber, 1991). From the first phase of the project the following lessons can be drawn:

- The development of performance measures requires data gathered on a uniform basis, which takes much time.
- Feedback of information to lower levels of management is a necessary condition for obtaining acceptance of the performance measurement system at those levels.
- Some organisational conditions may enhance the successful development of measures and ratios: these conditions include the creation of autonomous agencies such as regional inspectorates, with contract management for the supply of specific goods or services.
In spring 1991 a second phase of the pilot project was started. By then the climate surrounding performance measurement had changed. Former sceptics were now prepared to join the pilots and the initial propagandists had become more realistic. The purpose of the second phase was threefold: involvement of all departments to further develop ongoing projects (emphasising the use of performance measurement), making political use of performance measurement and improving the measures in the explanatory memorandum.

Two categories of pilots can be identified: one dealing with different types of autonomous government bodies and a second dealing with relevant aspects of implementation and developments in policy areas.

The first category focused on the demand for information on outputs by the parent departments for steering purposes. An important matter here was to prevent perverse reactions by the autonomous bodies such as achieving targets by reducing quality.

A lesson learned from these pilots was that steering on outputs is not a mechanical decision process but rather a process of periodic consultation. A notable point was also that in most pilots no explicit steering-model existed at the start of the experiment. The model was developed step by step in on-going discussions between all stake-holders.

In the second category of pilots some methodological problems had to be solved. The most important one was how to measure the efficiency of an organisation that produces different outputs. The problem is one of aggregating outputs that are physically or otherwise different, which could be solved by using total costs or time spent as weights.

An important lesson of these pilots was also that it is not easy to find comparable organisations for assessing the efficiency of one’s own organisation; and even if the organisations are comparable it is equally difficult to draw conclusions about differences in efficiency between these organisations. In most cases there are good reasons for these differences.

In 1992 and 1993 the third phase of the pilot-project was conducted. In this phase the attention was focused on the efficiency of the implementation of major policy programmes.

The approach involved four steps:

♦ providing a concise description of the various elements of the production process (input, activities, output, and effects), preferably subdivided into relevant economic subcategories (e.g. personnel and goods);

♦ developing adequate efficiency measures; the options are:
  ◊ output/input measures (unit-costs); and
  ◊ measures showing the relationship between programme management costs and programme costs;

♦ comparing efficiency measures over time with other organisations (cross sectional analysis), or with specific political promises;

♦ drawing conclusions for improving the efficiency of the production process.
These studies have yielded several interesting points. A major aspect has to do with programme management costs. Two approaches are distinguishable:

♦ Focusing on the organisation: In this case the relationship between the management costs of the organisation and the various policy programmes (outputs) are considered.

♦ Focusing on a specific policy programme: In this case the relations between the specific policy programme (outputs) and the costs of various responsible organisations are considered. Depending on the information requirements more cost categories are taken into account (e.g. the costs of policy development).

Another aspect has to do with the usefulness of the relevant comparison of performance measures. At a “lower” level (e.g. comparing efficiency of regional agencies) cross-sectional analysis may be useful. In most cases comparison over time is the right approach.

In evaluating efficiency quality aspects should also be taken into account. Increased productivity as a result of lower quality is hard to interpret so separate quality measures have to be added to pure efficiency measures.

An important conclusion of this study is that technical/methodological problems in general are solvable. So the specific character of the Netherlands central government administration (cash basis of accounting, no cost-price calculations) does not prevent obtaining data on efficiency. This implies, however, that in some cases in addition to the regular information on cash costs that estimates of depreciations are necessary. In other cases the costs of policy programmes can be determined in an indirect way on the basis of recording the time spent on activities. As shown in the second phase of the pilot projects this approach can also be used to determine the aggregated product of an organisation that produces various products (policy programmes) (De Groot et al., 1993).

Initiatives by the Ministry of Home Affairs

To enhance quality of public management the Ministry of the Home Affairs initiated the project “Kerngegevens Openbaar Bestuur” (Key figures public administration) of the Institute for Research of Government Expenditures (IOO) and the Social and Cultural Planning Office (SCP). The main purpose of this project, which was conducted from 1992 till 1995, was to transform published information from departments and already existing data at the IOO and SCP into usable efficiency-measures of the public sector and the non-profit sector receiving funding from the public sector. In this project progress was also made in solving some methodological issues, the most important of which is how to measure the quality of output of services.

The main conclusion of this project was that available information for measuring efficiency is not fully used.

b) Interdepartmental co-ordination

An important agent in the implementation of performance measurement was the creation of the interdepartmental support group on performance measurement, set up by the Ministry of Finance in 1990. In the early years this group consisted of a small number of enthusiastic directors in the field of financial management. Later the group expanded to include all departments. The primary task of this group, chaired by the Ministry of Finance, is to stimulate performance measurement in departments.
Pilot projects are discussed as are further developments in the value of measures over time. The Directorate of Budget Affairs initiates this and reports on results.

The main elements of these developments are:

- The rules for the departments for reporting performance measures. The departments should:
  - make explicit, in each budget item, the specific goal of the reported performance measures — insight into budget estimates, insight into efficiency or insight into effectiveness;
  - report the conclusions on relevant performance indicated by these measures; and
  - indicate explicitly the expenditures involved.

- The preparation of a manual with guidelines for departments at the end of 1994, translating the fourth amendment to the Government Accounts Act into operational terms. The results of the pilot-projects were used for this manual. The budget regulations will be readjusted in this respect. In the manual different categories of measures are identified — measures for giving insight into budget estimates, into efficiency and into effectiveness. Moreover the individual measures that are essential under each category are specified;

- The obligation of the departments to give plausible explanations for each category of measures or why certain budget items cannot be explained in a meaningful way with performance measures. On the basis of this information bilateral discussions between the Ministry of Finance and the other departments take place to reach an agreement on what is acceptable to both.

An overview in the budget memorandum of the Ministry of Finance of the percentage of each department’s budget that can be explained in a meaningful way by performance measures. Total expenditure of a department is not the point of reference here but rather the possibilities of explaining budget items with performance measures. The Budget Memorandum 1996 presented for the first time a detailed overview of these percentages for measures that give insight into budget estimates (see Chart 1 next page).
In 1997 a comparable overview will be presented on output measures.

An important initiative in interdepartmental co-ordination was laid down in the memorandum on policy evaluation in the civil service, sent by the Minister of Finance to other ministers in August 1994. In this memorandum attention was drawn to the agreement reached between the Interdepartmental Committee of Budget Directors (IOFEZ), representatives of the Netherlands Court of Audit and the Ministry of Finance on the specific character and quality criteria for policy
evaluation studies and the responsibilities of the various participants involved in policy evaluation (Ministerie van Financien, 1994).

This memorandum originated in the Official Policy Statement on Evaluation. The most important statement has to do with the primary responsibility of every minister for his own evaluation (studies) (Tweede Kamer der Staten- Generaal, 1991-1992).

On the first topic the memorandum offers a frame of reference for policy evaluation studies. Three categories are distinguished. The first category concerns evaluation in the form of systems of statistical measures that provide periodic insight into the Government’s policy performance. According to the memorandum this category is aimed at monitoring the implementation of policy for planning and control cycle purposes; sometimes these systems are integrated into the normal budget process. The second category deals with ex ante and/or ex post evaluations of effectiveness of specific policy programmes. The third category concerns organisational reviews (including management audits).

On the second topic it is stated that:

- policy directorates of the various departments have primary responsibility for their own evaluation, and thus also for their own performance measurement;
- the budget director of each department has a special responsibility for the planning and co-ordination of the evaluation activities;
- the Ministry of Finance (Directorate-General of the Budget) has the same responsibilities for promoting and co-ordinating the evaluation activities of other ministers as the budget directorates of each ministry compared to the policy directorates of a ministry;
- every year both the Court of Audit and the Ministry of Finance will be informed on the evaluation activities of the ministries.

c) Training

For many years the Ministry of Finance’s interdepartmental training courses on policy analysis and financial management included some modules on performance measurement. Until the end of the 1980s these were focused on the required information on budget items. Parallel to the development of a conceptual framework for performance measurement at the end of the 1980s courses on performance measurement in the field of efficiency and effectiveness were provided. In this way the necessary knowledge for good performance measurement was disseminated.

Political Support

It was partly due to parliamentary interest that in the 1992-1993 budget year pilot studies on the efficiency of the implementation of major policy programmes were conducted (Kabel et al., 1992). The general objective was to get information on the pros and cons of measurement systems especially on current and future issues of productivity and efficiency both for politicians and public managers. In addition the Ministry of Finance started an interdepartmental study on efficiency aspects of a number of major policy programmes. Nearly all departments participated in this project.
The main conclusion of these pilot studies was that apart from bureaucratic problems, both management and politicians can use relevant information on the efficiency aspects of policy programmes. For this reason, in the fall of 1994 the report on this study was sent to Parliament (Ministry of Finance, 1994).

In 1995 parliament took the initiative to study the possibilities of implementing specific efficiency measures in the explanatory memorandums of the budget. In the reports on the budgets of three departments made by consultants, recommendations were made for the appropriate efficiency measures at each budget-entry of the explanatory memorandum.

**Agencies**

Until recently all government organisations (except state enterprises, which since 1 July 1994 no longer exist) were subject to the same management rules.

In 1991 the government decided to relax the general rules governing finance and personnel in various ways. It was also recognised that some services could benefit from a different management system. For these services a new type of fairly autonomous government body has been created: the agency, a concept borrowed from the United Kingdom.

Four agencies were set up in 1994, three agencies followed in 1995 and six agencies will follow in 1996. Agencies are required to use an accrual accounting system rather than a commitment/cash system.

There are several conditions which agencies must meet. The management of an agency focuses on achieving objectives and not on spending or inputs. An agency is required to account explicitly for the production level or outputs achieved. This means that production must be measurable. It must be shown that an agency is unable to operate on the basis of “normal” government rules and can operate more efficiently with a tailor-made management system. Its administrative organisation must also function properly. An unqualified auditor’s report is also a precondition for a service becoming an agency.

Since an agency involves a form of commercialised government organisation it goes without saying that before a service becomes an agency, it must show that it fulfils at least those conditions. This takes a quite long period of time (say at least a year). In this period there is a lot of communication between the parent ministry, the service itself and the Ministry of Finance. In the end the Minister of Finance will present to parliament the proposal that the service in question become an agency.

One of the features of an agency is that management is based on output rather than input. Performance measures are necessary to make agreements with an agency on the production level to be achieved as well as on the way that the agency will show that it operates more efficiently. This information should be provided in the explanatory notes to the budget provided to parliament. The report sent to parliament at the end of a year should show whether the agency has been able to carry this extra responsibility and use its resources in such a way that the budgeted level of production or the intended improvement in services has been achieved.

It is very important that the right measures — that is, those which an agency uses for management purposes — are determined and later reported on. The management of departments that
become agencies should be organised according to the new philosophy. This is not something that can be done overnight.

In practice two main problems have to be solved. First is the definition of the products or outputs. Most of the agencies have some activities which cannot be translated into products. For example the Royal Netherlands Meteorological Institute (KNMI), which falls under the Ministry of Transport, Public Works and Water Management, does, besides the weather forecasts, a lot of basic meteorological and seismological research. Defining the outputs of those activities is difficult. In this case the ability to measure production is limited to process-measures. The second problem to be solved concerns accounting for the costs of the products (or process indicators). In most ministerial departments it is not common to allocate costs to products; even the time spent on the different products is in most cases not known. So besides solving the accounting problem, a cultural change is needed.

There is discussion between the parties on the problems mentioned above. Gradually the problems are being solved and measures created which give insight into the position and the progress of the agencies.

Evaluation and Recommendations

The general conclusion is that performance measurement is technically possible. The efforts involved are relatively simple in comparison to sophisticated ex ante or ex post evaluation studies (Tweede Kamer der Staten-Generaal, 1991-1992). Moreover there is no need for a drastic change in the budget system. Where it is shown to be necessary the agency status is a suitable solution.

On the question of what approach should be followed to further stimulate performance measurement, a strategy in which some pressure is used by the co-ordinating agency may offer the best chance of success.

A favourable circumstance is that the present interest of Parliament and management offers the right climate. In such a climate all parties involved are committed to overcome bureaucratic and methodological problems in performance measurement.
NOTE

1. In this paper “measures” is used as a synonym for “measures and/or indicators”.
REFERENCES


Chapter 7

DEVELOPING AND USING PERFORMANCE MEASUREMENT: THE NETHERLANDS EXPERIENCE

Commentary by

Steve Nicklen
United Kingdom Audit Commission

I appreciate the opportunity to comment on the most interesting Netherlands’ paper. Our experience in the United Kingdom Audit Commission reinforces many of their conclusions but there are also some interesting differences between their and our approaches.

I should say just a few words of background. The Audit Commission determines which performance measures UK local authorities should use to monitor the services they provide (essentially education, personal social services, highways and planning, police, public housing, but also a range of others, involving well in excess of $US100 billion of public expenditure each year). Local authorities must then collect information in relation to those performance measures and publish their performance in a newspaper. The Audit Commission then publishes national comparisons of local authority performance based on these performance measures.

The first Directive setting out these performance measures was issued by the Audit Commission in 1992. Local authorities then collected information for 1993/94 and published it in local newspapers during the second half of 1994. The Audit Commission then published nationally for the first time early this year. Major press coverage resulted. All quality newspapers and almost all local newspaper and radio stations (some 300 in total) reported this.

Our experience supports that of the Netherlands in a number of areas:

♦ First, the importance of analysing the use of performance measures and learning from that experience. We experienced a lot of scepticism from local government politicians, most of whom resented what they saw as an imposition from central government, especially when that central government is usually under a different party political control. We have put a lot of time into researching how the performance measures are used and in particular how the public responded to the first year of national publication. The evidence is clear. Both the press and the public welcomed these performance measures and their publication. And this, in turn, has done a lot to silence our critics.

♦ Second, we also set ourselves the twin objectives of both improving accountability (to Parliament in the case of the Netherlands experience, to local government politicians and to
the public they serve in our case) and also to foster the efficiency of service delivery. Like the Netherlands we had an explicit objective in our work to bring about improvements in the economy, efficiency and effectiveness of service delivery, not just to report on them.

Third, the lack of interest by politicians. Candidly, this has been a major problem for us to date. Most UK local government politicians, and all too many managers was well, have seen the indicators as something to react to defensively rather then as an opportunity for improvement. So we are now focusing a lot of our effort on trying to get politicians more involved. We are running conferences around the country for politicians and we are producing publications about the performance measures aimed specifically at politicians. But I do not underestimate the difficulty of making progress here. There is a cultural problem.

Fourth, the need to look at performance in all its aspects - at inputs, efficiency and effectiveness. We, like the Netherlands, have found it much easier to deal with the first two, than the last. We have found it hard to measure effectiveness, let alone net-effectiveness, of particular services. In approaching the general issue of quality we also are attracted by the idea of introducing consumer satisfaction surveys. But we are moving only tentatively down this route because of the cost of carrying out surveys.

Fifth, the need for comparisons to be made over time, with other organisations, and of actual performance against standards. In our first year of national publication we stressed comparisons between organisations. But in doing this we also pointed out that there are huge differences in local circumstances between different local authorities which need to be taken into account in making such comparisons, eg in local geographical, socio-economic and demographic circumstances. Early next year for the first time, in the second year of national publication, we will be seeking to encourage comparisons over time. And early next year we will be publishing a management paper trying to spread best practice on how to set local standards as a basis for judging performance against performance standards.

Sixth, the need to provide explanatory memoranda in support of the performance measures. Definitional problems about what particular performance measures mean have been a major issue for us. We have worked hard with professionals in local authorities to provide tight, unambiguous definitions. And we operate a central helpline at the Audit Commission which local authorities can contact to help them interpret these definitions. This has helped enormously to reduce the problem of not comparing like with like between different local authorities. I believe it has been crucial to our success to date. But I would not pretend that it has yet eliminated the problem of definitions.

But there are two areas in which I think our experience is different from that set out in the Netherlands’ paper:

First, we have shied away from trying to measure the total efficiency of an organisation that produces different outputs. The problem of aggregated outputs, as the Netherlands paper describes it, or producing composite performance measures, as we refer to it in the Audit Commission, has so far been something to which we have not been able see a technical solution.

As in many countries, the approach was to provide an inter-departmental support group on performance measures, but to look to individual Government agencies to develop their own
performance measures in a bottom-up way. Ours was fundamentally different. We did consult with professionals and managers, very extensively and widely. But at the end of the day the Audit Commission came to its own view and imposed standard performance measures in a top-down fashion.

So to summarise, the Audit Commission’s experience in the United Kingdom in developing performance measures for local government services supports the Netherlands’ experience in most important respects, but with a couple of important differences.

Important themes which flow from the Netherlands paper are:

♦ How do we get politicians interested in what performance measures show?

♦ What are the alternative to consumer satisfaction surveys in looking at the quality of service delivery?

♦ How can we approach the issue of aggregated outputs/composite indicators?

♦ What are the benefits and disadvantages of “top-down” and “bottom-up” approaches?
PERFORMANCE MANAGEMENT IN GOVERNMENT
CONTEMPORARY ILLUSTRATIONS

The Public Management Occasional Papers are specialised reports prepared for the work of the OECD’s Public Management Committee.

Two key related elements in public management reforms in OECD Member countries are greater devolution of decision-making authority in government organisations and an increased emphasis on their performance. Together they constitute what is frequently described as performance management.

The papers in this Occasional Paper illustrate and analyse important developments in performance management in a range of OECD countries, with particular emphasis on issues in developing performance measures in government organisations.