WE31 Improving evaluative activity in the New Zealand state sector

A paper presented at the Australasian Evaluation Society Conference, 16 – 18 September 2003, Auckland, New Zealand

Shenagh Gleisner,¹ Arati Mascarenhas² and Sonia Wansbrough²

Abstract

This paper outlines the way in which the New Zealand Government is intending to improve the targeting and use of evaluative activity in the state sector.

In 2001, the government made a commitment to improve the public management system in New Zealand. One of the recommendations required officials to consider what could be done to improve the "evaluation environment" and to encourage more effective use of evaluation in New Zealand. The project team analysed the strengths and weaknesses of evaluative inquiry in the New Zealand state sector - how it is targeted, conducted and used. The team found many examples of good practice across the state sector. Recent initiatives are likely to further improve the evaluation environment in the medium to longer term. There were, however, some enduring problems. The project team recommended changes to the public management system to enhance evaluative activity, support for developing evaluative capability and a process to coordinate evaluations of major policies.³ In August 2003, Cabinet agreed to these recommendations.

Introduction

A project team from The Treasury and the State Services Commission (SSC),⁴ with guidance from an advisory panel of evaluation, policy and statistics experts, has looked at the current state of evaluative activity in the New Zealand state sector, recent initiatives that are improving the situation, and what is needed to build on improvements and address any remaining problems. Cabinet agreed to the recommendations arising from the project in August 2003. This paper outlines the findings of the project, its recommendations and the reasoning underpinning the New Zealand Government's proposed direction.

¹ State Services Commission

² The Treasury

³ The final project report is at: <u>http://www.treasury.govt.nz/evaluation/</u> and <u>www.ssc.govt.nz/doing-the-right-things</u>

⁴ The State Services Commission (SSC) is the department that supports the State Services Commissioner in his role in maintaining New Zealand's politically neutral, professional and permanent Public Service. In accordance with the State Sector Act 1988, the State Services Commissioner appoints, employs and reviews the performance of departmental chief executives, sets standards of conduct and integrity for the Public Service and investigates and reports on matters relating to departmental performance.

Why we looked at evaluative activity in the New Zealand state sector

In 2001, the Review of the Centre (RoC) was undertaken to review the operation of the public management system in New Zealand and to recommend improvements. As part of its analysis, the RoC report expressed concern that evaluation is not conducted and used effectively enough in the New Zealand state sector. RoC was primarily interested in impact evaluation as a result of concern expressed in a number of earlier reviews that New Zealand did relatively little impact evaluation compared to other jurisdictions, especially Australia. RoC noted that there were developments under way – including the Social Policy Evaluation and Research Committee (SPEaR), the Pathfinder project⁵ and the greater outcomes focus in Statements of Intent⁶ – that might improve the evaluation situation. However, at the time, not much was known centrally about the targeting and use of evaluation because of the extent of devolution in the NZ state sector.

The RoC report therefore recommended, and Government agreed,⁷ that there should be a project to determine the likely impact of these developments on the targeting and use of evaluation, and whether additional steps might be necessary to achieve further improvement.

In New Zealand, there has not been a coordinated whole-of-government approach towards evaluation. Instead, decisions about what evaluative activity to undertake are generally at the discretion of individual agencies, and often with individual managers within those agencies. Individual departments have gradually developed their own capacity to conduct evaluations by developing evaluation and research units and their own evaluation strategies. However, without strong support and championing from the central agencies, ⁸ evaluative activity and capability is limited compared with other jurisdictions. Improvements in the frequency and effectiveness with which evaluative activity is conducted and used have occurred recently. Initiatives already under way will increase the focus on thinking about, and doing, what works best to achieve outcomes, improve coordination and sharing of information and data, and develop evaluative capability.

⁵ The Pathfinder Project is an initiative seeking to better integrate outcome information into the public sector management system. The project is a 'hot house' in which eight New Zealand government agencies are developing outcome management systems that meet their individual needs. For more information see: http://io.ssc.govt.nz/pathfinder/

⁶The Statement of Intent (SOI) is a public document, which provides transparent information to Parliament and the New Zealand public about how the department intends to manage for outcomes. SOIs have a medium-term (three-year plus) focus and aim to improve alignment with the Government's objectives through better planning for the contribution of outputs to outcomes and for capability. SOIs include the annual statements required by s.34A of the Public Finance Act.

⁷ In 2001, following the Review of the Centre, Cabinet asked for a project to assess whether initiatives, in addition to the establishment of the Social Policy Evaluation and Research Committee, would be necessary to enhance the evaluation environment and to encourage more effective use of evaluation in the state sector (CAB Min (01) 39/14 refers).

⁸ Central agencies are Treasury, SSC and Department of Prime Minister and Cabinet.

The project team expanded the scope of the project from looking at "evaluation" to consider the broader concept of "evaluative activity", since evaluations are not the only way to provide useful information about the efficiency and effectiveness of policies and programmes. We defined "evaluative activity" as activity that seeks evidence of actual performance of policies or programmes once they are being, or have been, implemented. In this sense, evaluative activity includes formal evaluations of policies and programmes, monitoring and performance audit.

Moreover, since evaluative resources in New Zealand are very limited, departments will need to make careful decisions about when to undertake evaluations and when other forms of evaluative activity will provide sufficient information at lower cost. Therefore, our analysis has considered how well officials decide what to evaluate, when to commission evaluative activity and what type of evaluative activity to undertake.

A central theme of the RoC report was the need to focus the public management system increasingly upon achieving outcomes for citizens. Consequently, our interest in evaluative activity is about how it can help improve decision-making, so that government interventions achieve better outcomes for citizens. We have therefore also looked at how effectively departments use the results of evaluative activity to inform policy and programme design and delivery.

What we did

We started our investigation by looking at previous reviews of the state of evaluation and relevant literature, including reports of developments in overseas jurisdictions. To ascertain the current state of evaluative activity in the New Zealand state sector, we tested the main themes and conclusions from the literature and previous reviews through interviews with policy managers, evaluators and academics, and comments from state sector agencies and other stakeholders on our findings.

We have not undertaken a full survey of evaluative activity and compared the current level with past activity in order to ascertain the change taking place. Nor have we quantified the amount of resource spent on evaluations within the State sector. We were advised by our advisory panel from the outset that such an exercise would not be feasible or accurate enough to provide much insight within the project timeframe. We have focused on what more needs to be done to build on existing good practice around the State sector rather than an in-depth analysis of what has driven developments to date.

What we found

Our main conclusions are that Ministers and officials increasingly use evidence obtained from evaluative activity to inform decisions about which policies and programmes are most likely to achieve sustainable results for New Zealanders in the long term and how best to deliver these. Improvements in the way that agencies decide what evaluative activity to undertake and the quality of evaluative activity are apparent and existing initiatives are likely to address some of the remaining problems. However, we found inconsistency across the state sector in how well agencies target and use evaluative activity. This section outlines our findings in more detail.

Findings of previous reviews

In the last five years, several reviews of the state of evaluation have been undertaken.⁹ These reviews have focused primarily on the extent to which impact evaluations are undertaken and used in policy. We started our investigation by looking at these reviews and relevant literature, including reports of developments in overseas jurisdictions.

These previous reviews as well as relevant literature found that:

- some of the essential prerequisites for good evaluative activity are not in place (e.g. strong and consistent interest in using evaluative findings);
- evaluative results, once generated, are often not well integrated into the policy process;
- insufficient evaluation of the impacts of policies or programmes is undertaken; and
- skills to produce robust information about the efficiency and effectiveness of policies and programmes, especially about interventions for Māori, are not well developed.

Good practice but not widespread and consistent

Contrary to many of the messages in the literature and previous reviews, we heard many good examples of Ministers and officials seeking and using information about efficiency and effectiveness to inform policy and programme delivery during our interviews and in response to our various papers. A number of departments have:

- good prioritisation processes to decide what to evaluate and what type of evaluative activity is the best use of available resources;
- strong links between evaluators, policy and operational staff;
- joint evaluation strategies with other agencies;
- processes to plan evaluations early when policies and programmes are being developed;
- developed and used culturally appropriate methodologies;
- involved stakeholders in design of evaluations; and

http://www.oag.govt.nz/HomePageFolders/Publications/1stReport2000/1stReport2000.htm; Office of the Controller and Auditor-General Review (1999), *Third Report for 1999:The Accountability of Executive Government to Parliament*,

http://www.oag.govt.nz/HomePageFolders/AuditOfficeReports/3rdReport1999/3rdReport1999.htm

⁹ State Services Commission, (June 1999a). *Looping the Loop: outcome evaluation and other risky feats*, SSC Occasional Paper No. 7. http://www.ssc.govt.nz/display/document.asp?docid=1725; State Services Commission (June 1999b). *Essential Ingredients - Improving the Quality of Policy Advice*, Occasional Paper No. 9, http://www.ssc.govt.nz/display/document.asp?docid=1763&pageno=3#P96_11284; Office of the Controller and Auditor-General (2000) *Impact Evaluation – its Purpose and Use*,

• reported and disseminated results so that they can be easily understood by nonevaluators (including stakeholders).

However, this good practice is not widespread. In particular:

- the interest in commissioning evaluative activity and commitment to using evaluative findings is variable;
- understanding by policy and programme managers of when different types of evaluative activity offer value, how to interpret and use findings is limited;
- skills within and outside the state sector to conduct high quality evaluative activity are limited;
- evaluative effort is not always well prioritised and coordinated, so that major policies that cross agency boundaries are often not evaluated;
- sharing and consistency of data and evaluative findings within and between agencies is limited; and
- a lot of evaluative information about policies and programmes for Māori exists although it provides limited evidence about what those policies and programmes actually achieve for Māori.

Variable culture of inquiry

Some agencies have developed evaluation strategies to guide investment of evaluative resources and capability to effectively conduct and use evaluative activity because their Ministers or senior officials have seen it as important. These agencies have strong evaluative cultures – they use evaluative findings to inform policy, service delivery or broad government strategy and budget decision-making decisions. However, many agencies have weak evaluative cultures because:

- evaluative activity is not always requested as an input to decision-making;
- the incentives for evaluative activity throughout the public management system for example the emphasis put in chief executive performance management and the budget processes are not strong enough; and
- capability to effectively commission, conduct and use evaluative activity is limited.

Ministers make decisions in a political environment, which means that more factors come into play than just evidence of efficiency or effectiveness. As a result, they will sometimes make decisions to continue certain approaches even when there is evidence that they are not working. However, policy advice from officials should be "free and frank" and therefore based on the best available evidence.

The public management system does not strongly reinforce the need to base advice on the best available evidence. The actual performance of programmes and policies is subject to limited scrutiny from central agencies. Despite this, some agencies and Ministers do place a great emphasis on basing advice and decisions on evidence.

In some cases, Ministers and officials do not use evaluative findings because findings are not available in time, do not provide conclusive results or are presented in a form that is not accessible. This occurs either because of data and methodological constraints, limited capability to effectively commission, conduct and use evaluative activity amongst evaluators and policy staff and a disconnect between evaluative activity and the policy process at times.

Poor coordination and prioritisation across the state sector

Evaluative effort is not always well prioritised and coordinated, and sharing and consistency of data and evaluative findings within and between agencies is limited. This is because most decisions about what evaluative activity to undertake are at the discretion of individual agencies. However, few mechanisms exist to keep track of what evaluative activity has already been undertaken and to coordinate and prioritise future activity. Incentives to coordinate or prioritise are also relatively weak. Whilst the SSC is moving towards encouraging collaboration around outcomes, there is still a strong focus upon a chief executive's own departmental performance.

The limited mechanisms and incentives for coordination and prioritisation appear to stem from a range of factors including transaction costs associated with coordination and the varying significance of a particular outcome for contributing agencies.

In particular, there are limited incentives for individual contributing agencies to coordinate evaluation around policies to which multiple agencies contribute. Therefore, many substantial, expensive policies have not been adequately tested for effectiveness and efficiency. This applies to policies and programmes that fall within the scope of individual agencies but is exacerbated for those that rely on multiple agencies. Many of the Government's priority outcomes rely on interventions from multiple agencies. Some agencies do work together to evaluate policies that span agency boundaries but in most cases, collaboration is limited.

In some sectors, lead agencies drive coordination and prioritisation between related agencies. For example, the Ministry of Social Development and the Ministry of Justice play this role to some extent. However, lead agency mandate and scope is unclear, so this approach is not yet widespread and not as complete as it could be. Another recent improvement in the social sector is SPEaR's mapping project, which collates the research and evaluation programme activities of some 20 agencies.

Limited capability to effectively commission, conduct and use evaluative activity

Evaluative skills within both the State sector and the private sector to conduct high quality evaluative activity are relatively limited compared to existing demand. The fragmented nature of the State sector may also contribute to evaluative capability shortages by spreading it thinly. Finding peers to review work, given the limited community of practice, can also be hard.

A major obstacle to improving evaluative capability in the State sector is the lack of specific evaluation training and development opportunities. Many evaluators must train

overseas, receive on-the-job development where it can be fitted in (placing further pressure on the small number of more experienced evaluators), or simply fend for themselves. Tertiary institutions have recently started to increase their focus on evaluative training. The fact that there is a level of evaluative expertise in the State sector, however, reflects deliberate efforts to build capability by chief executives and the Australasian Evaluation Society and local evaluation societies.

As mentioned earlier, limited capability amongst those who commission, conduct and use evaluative activity sometimes limits the usefulness of evaluative findings. In order to make good decisions about what evaluative activity to undertake, advisors and decisionmakers need to understand what different types of evaluative activity offer at what cost and within what timeframe. However, people without this understanding often plan evaluative activity, without input from evaluators, resulting in poor decisions about what to evaluate.

In addition, many respondents expressed concern that policy staff generally have relatively poor evaluative and statistical analysis skills. This limits their ability to understand and use evaluative findings effectively to inform policy and programme design. In particular, many policy staff reportedly expect too much certainty from evaluative findings and discredit evaluations without conclusive results.

The technical language in which evaluative findings are often presented also contributes to poor understanding among actual and potential users of these findings. Findings are often not communicated clearly and simply so that they can be understood by people without technical evaluation expertise. This may be the result of some evaluators not having the skills to communicate complex technical findings to a range of non-technical audiences.

The state of evaluative activity for Maori

Much evaluative information about policies and programmes involving Māori exists. Te Puni Kokiri¹⁰ reviews of departments' policy and programme evaluations show some good practice, including some development of guidance specific to evaluation of programmes involving Māori.¹¹ At times, evaluative activity of programmes involving Māori draws on culturally appropriate methodologies.

However, evaluative activity often provides limited evidence about what the interventions actually achieve in improving outcomes for Māori. Moreover, consultation processes with iwi and hapu vary by agency and by evaluation projects. Some agencies do not engage with Māori stakeholders appropriately. Often Māori do not receive results of evaluative activity or see any improvements to programmes or to their communities as a result of evaluative activity.

Many factors contribute to these problems, including:

¹⁰ Ministry of Māori Development

¹¹ Te Puni Kokiri, *Review of the Social Policy Agency, Te Ropu Here Kaupapa*, Aroturuki me te Arotakenga, Monitoring and Evaluation Branch, August 1999, 53.

- a limited supply of evaluators with an adequate understanding of methodologies appropriate for Māori as well as sound evaluative skills and acceptability in Māori communities; and
- officials commissioning and using evaluative activity to inform policy often do not know how to measure the effectiveness of the policies or programmes aimed at improving outcomes for Māori.

Initiatives underway that will improve the situation

This section outlines some of the recent initiatives intended to focus departmental effort on achieving outcomes, improving coordination and developing capability. To some extent these will address the problems noted above.

Managing for Outcomes and Pathfinder are intended to encourage better planning and management by departments. This includes a focus on reviewing performance and using evaluative information to modify the activities that they undertake to enhance their effectiveness. Therefore, these initiatives are emphasising the need for information about what outcomes are being achieved, what part government interventions have played in their achievement, as well as how interventions could be changed or improved so as to increase their impact.

The Social Policy Evaluation and Research Committee (SPEaR) was set up in 2001 to oversee the government's social policy research and evaluation purchase to ensure the spending is aligned with the Government's social policy priorities. This group of evaluators and researchers is driving improvements in social evaluation and research capacity and capability, and prioritisation and coordination across the social sector. For example, in addition to the mapping project, SPEaR is developing best practice guidelines, including guidelines for the evaluation of programmes and policies involving Māori,¹² and facilitating joint training and development of personnel in relation to evaluative activity. The impact of SPEaR's programme will be clearer in mid 2004 when its progress is reviewed.

Statistics New Zealand's administrative data integration project will help synthesise data across the government sector. The current review of official statistics is likely to recommend some changes to the way that administrative data are collected and the place of official statistics. Such changes have the potential to improve the quality of information collected about policies and programmes and coordination between agencies.

The newly established Australia and New Zealand School of Government is building policy, research and management capability, and other tertiary institutions are offering more courses relevant to evaluative activity.¹³ The Executive Leadership Programme (as

¹² These guidelines are intended to supplement the existing Te Puni Kokiri guidelines.

¹³ For example, Massey University now offers a postgraduate diploma in evaluation.

part of the Senior Leadership Management Development Programme) will provide future Public Service leaders with an appreciation of the importance and value of evaluative thinking and activity.

What's now happening as a result of our project

The project team recommended several changes, which were all supported by our Ministers and then accepted by Cabinet in August 2003. We considered a wide variety of options including separate evaluation funding streams, mandatory evaluation for new initiatives and a centralised evaluation unit. However, our main recommendations focus on getting departments to consider evaluative activity as part of their normal business and to make the appropriate decisions for their organisations about capability and the targeting and use of evaluative activity. Our approach builds on the already positive developments in New Zealand.

The primary initiatives arising from this work include:

- emphasising the importance of reviewing performance as well as planning through Managing for Outcomes guidance and training;
- providing some training and guidance about the different types of evaluative activity and appropriate uses of each type of evaluative activity through Managing for Outcomes;
- supporting departments as they develop evaluative capability through a community of practice;
- advising the Ministers of central agencies about areas in which coordinated evaluative effort will provide the Government with information about major policy areas that they would not otherwise receive; and
- supporting the targeting and use of evaluative activity through the budget process.

As part of the Managing for Outcomes initiative, central agencies are emphasising the importance of undertaking prioritised evaluative activity. Departments will be expected to develop an evaluation strategy that prioritises their evaluative activities within the context of their overall policy direction and that of related agencies. Managing for Outcomes training will cover the value of different types of evaluative activity at different stages of the policy/programme cycle, good information management practices and understanding and using evaluative findings.

SSC will establish an evaluation and monitoring community of practice, including private sector and Māori evaluators, to share evaluative good practice and address capability needs. The community of practice will share information about training opportunities, and may also deliver seminars or training workshops. The community of practice will bring together evaluation practitioners and a range of universities, including the Australia and New Zealand School of Government.

Treasury, SSC and the Department of Prime Minister and Cabinet are currently designing a process to coordinate advice to Ministers about where evaluative activity could be undertaken to help inform their decisions about the future of major policy areas. Since major policy areas often cross agency boundaries, central agencies will play a coordinating role. This will include consulting with departments and relevant organisations on existing or planned evaluative activity and what additional evaluative activity would add value.

Treasury is currently changing the budget process so that it is more focused on overall government expenditure. Departments will be encouraged to use the evaluation strategies developed as part of Managing for Outcomes to guide their decisions about future evaluative activity and to use evaluative findings to inform new initiative proposals.

Why this direction?

A number of approaches were possible to respond to the direction by Cabinet to enhance the evaluation environment and improve the use of evaluation in the New Zealand state sector. Our approach was informed by three key factors including the problem definition, the levers available in the public management system and a desire for consistency with the style and direction of the current government.

Primary focus upon enhancing demand for evaluative activity

Our research suggested that the most pervasive problems were in the commissioning and use of evaluative activity. So, in our recommendations, we have focused on building demand for evaluative activity. The quality of evaluative activity itself is high in many parts of the public service although it is not consistently high. Unless decision-makers commission evaluative activity wisely and evaluative activity is well designed to provide assistance to decision-making, evaluative findings are unlikely to be useful or used. Evaluators frequently commented that evaluation was commissioned to provide "answers" which were not feasible in the timeframe or that evaluation reports were not consistently well used. To address these problems, we have focused primarily on building demand for evaluative activity rather than building the supply of evaluators.

We are primarily interested in improving capability to commission evaluative activity and use evaluative results. So, we are aiming, in the first instance, to improve the understanding and application of evaluative thinking in management and amongst decision-makers. Our emphasis upon capability is in the following areas:

- creating a culture of enquiry attempting to enhance the focus and skills of leaders on how to commission and use evaluative findings to inform decisions;
- prioritising evaluative effort encouraging analysis and planning which ensures evaluative activity is focused upon major strategies and interventions;
- educating policy and programme managers people tasked with design and delivery must know when and how to use evaluative activity and how to interpret and use results; and
- coordinating efforts including the sharing of data, methods and findings.

We are conscious that it will take time to develop sufficient capability to meet increasing demand. Therefore, we have recommended some initiatives to build evaluative capacity and the skills to conduct high quality evaluative activity.

An evolutionary approach, seeking cultural change not compliance

Another key decision that we made was to recommend a voluntary approach, rather than mandatory requirements.¹⁴ Instead of promoting a compliance approach towards evaluative activity, the New Zealand government hopes to build a genuine culture of inquiry.¹⁵ This focus on changing culture is consistent with other initiatives underway such as Managing for Outcomes and a change in the way that central agencies are interacting with departments. For example, previously SSC focused on specifying requirements and assessing departments against those requirements. Assisting departments was a lesser focus. SSC and other central agencies are now focusing more on leadership, brokerage, facilitation and innovation. Within this climate, mandatory requirements are less appropriate.

Furthermore, we considered that mandatory requirements would result in exactly the problems we were seeking to avoid. It could lead to inadequate evaluative skills being applied to marginal programmes rather than to the programmes or questions most likely to provide useful and valuable findings to inform future decisions. Because resources will always be scarce, the most important thing is that evaluative activity is applied to the areas where greatest gains can be made. This may not be within departmental boundaries – it may be most importantly applied to inform government on outcome areas that cross agency boundaries.

Using incentives and levers through the public management system

Our project focused on how the public management system could be adjusted to enhance the evaluation environment. We adopted this approach to be consistent with the Review of the Centre and to embed evaluative thinking into the every day management of agencies. The public management system provides the context within which chief executives manage. Departmental chief executives are given freedom to manage their departments but are held accountable for delivering outputs and managing for outcomes using available resources. Central agencies help operate the levers of the public management system, specifically the chief executive performance management process, strategy setting and the resource allocation process. Central agencies can make certain

¹⁴ By contrast, at the outset, evaluation policy in Canada and Australia was centrally mandated and compulsory. However, changes during the 1990s in Canada and late 1990s in Australia have led to a more decentralised system in both countries.

¹⁵ For example, in reference to the evaluation requirements introduced in Australia in the late 1980s, the Task Force on Management Improvement found that even though "the evaluation strategy is intended to encourage a focus on results . . ., ironically, there may be a tendency still to focus on satisfying the requirements for evaluation (that is, observing the letter of the strategy rather than learning to use evaluation to improve program outcomes). (Task Force on Management Improvement, *The Australian Public Service Reformed: an evaluation of a decade of management reform*, 1992, p379)

behaviours more highly valued through their leadership and the incentives they apply through the public management system.

The project has therefore focused on how the levers of the public management system – for example, the budget process, chief executive performance management, strategic planning and annual reporting – could be modified to encourage and support departments to develop evaluative cultures.

What improvements do we expect to see?

Our recommendations will not improve the state of evaluative activity immediately. Changing culture, developing capability and improving information systems take time and resources. The Treasury and SSC will conduct a review in December 2005 to assess whether there has been an improvement in evaluative activity in the State sector and whether further work is necessary. At that time, we will reassess some of the decisions we have made about our proposed direction. This could include the decisions to focus on demand side issues and the public management system.

Nonetheless, we do expect to see some improvements in the way departments use evaluative activity. In particular, we hope to see:

- departments considering how to assess the efficiency and effectiveness of policies and programmes and involving evaluators and stakeholders as they design policies and programmes more often;
- departments using evidence from evaluative activity to inform policy and programme decisions and to inform budget initiative bids, including decisions to reprioritise existing activity and resources where it will be more effective than seeking new funding more often;
- greater awareness of evaluation approaches for programmes for Maori and use of Te Puni Kokiri and SPEaR guidelines for evaluating programmes for Maori as well as the use of culturally appropriate methodologies where necessary;
- improvements in the quality of information and more systems to provide and make use of such information, within and between agencies.
- enhanced understanding by all stakeholders of what different types of evaluative activity offer and what it takes to do them well;
- use of a variety of research and evaluative methods that fit the purpose of evaluative activity and the needs of different organisations; and
- a greater supply of personnel with the knowledge and skills to plan, conduct and use evaluative activity.