OCCASIONAL PAPER

No. 7

Looping the Loop: Evaluating Outcomes and Other Risky Feats

STATE SERVICES COMMISSION

Te Komihana O Ngā Tari Kāwanatanga

This paper was prepared as part of the State Services Commission's "Improving the Quality of Policy Advice" project, and examines the extent of outcome evaluation in the New Zealand Public Service. It identifies some of the systemic disincentives to evaluation, and suggests mechanisms for encouraging departments to develop a greater capacity for outcome evaluation.

June 1999

This paper was prepared in May 1998 as a contribution to the State Services Commission's project "Improving the Quality of Policy Advice". Since then changes have been introduced to the policy environment to achieve a greater focus on outcomes, including through the introduction of Ministerial Teams and wider adjustments to the strategic management framework. An updated version of the ideas presented in this paper, which takes account of those changes, appears in Occasional Paper No. 9 on policy inputs entitled "Essential Ingredients: Improving the Quality of Policy Advice".

Contents

To obtain a copy of this paper contact:

Ken Martin
Publications Distribution
State Service Commission
P O Box 329
Wellington, NZ

telephone: +64 4 472 5639
facsimile: +64 4 473 5979
cell-phone: 025 446 243
email: ken.martin@ssc.govt.nz

or

visit the Commission's website: http://www.ssc.govt.nz
and access the paper through the homepage search facility

For further information about the paper, or to make comment, please contact:

Derek Gill, Branch Manager
Strategic Development Branch
State Services Commission

Email: derek.gill@ssc.govt.nz   facsimile: +64 4 495 6699

Introduction

> *"Having moved away from inputs to output purchasing and all that entails, the challenge now is to focus on the outcomes the community and Government of the day seeks to deliver"*
>
> Rt. Hon Jenny Shipley
> March 1997

Numerous commentators have expressed concern that the focus on outputs in the New Zealand public management model (NZPMM) has meant that far too little attention has been paid to the outcomes or goals a government seeks to achieve. The strategic management system, including SRAs and KRAs, was put in place precisely to enhance the links between outputs and outcomes. Yet, despite this system, there is lingering disquiet – expressed explicitly by the then Minister of State Services, Paul East, in late 1997 – that departments might be simply 'best-fitting' their existing departmental outputs to convenient SRAs. That is, rather than driving KRAs and outputs, SRAs are used as justification for continuing business as usual.

On a day to day basis, broad outcomes seem to be even less evident as the driving force behind or raison d'être for policy proposals, and policy advice outputs. Departments appear to be able to identify the potential 'winners and losers' in a policy proposal – although sometimes even this advice is 'spotty' – but are less able or willing to cite the potential impact on strategic outcomes. Even more worrying is the fact that very few departments actually monitor, review, or evaluate the extent to which the policy outputs they produce contribute to government strategic priorities. Therefore, if ex-ante links between outputs and outcomes are weak, ex-post links are in an even worse state. The intervention logic might be spot-on; policy advice and resulting decisions might be having the desired effects (difficulties with establishing causality between policy advice and final results notwithstanding). But in the absence of some form of evaluation against outcomes we cannot say for sure. How can we tell whether outcomes are being met or even if we are getting closer to the target? Without evaluation, or other forms of ex-post review, the quality of policy advice remains unknown.

The purpose of the paper

This paper describes the extent to which departments are rising to the *evaluation challenge* and attempting to assess the impacts of their policy advice outputs against outcomes sought. This information draws on interviews with senior policy managers conducted as part of the State Service Commission's project on *Improving the Quality of Policy Advice* (IQPA). The views of those managers are expressed as quotations throughout the paper. The paper argues that evaluation is an important tool for examining links between outputs and outcomes, and attempts to define the necessary conditions for the development of a culture that encourages review and evaluation. The paper also looks at the extent to which review and evaluation exist in the New Zealand context, and argues that while the current environment is not overly conducive to outcome evaluation, there is nothing inherent in the public management model that inhibits it. There is room, however, to improve incentives to evaluate, and

some potential levers are proposed. The paper concludes that the difficulties associated with outcome evaluation should not deter public servants from attempting to 'loop the loop' by building review and evaluation into the policy advice process.

Definitions: evaluation

There are a multitude of definitions of evaluation. They range from narrow interpretations based on quasi-experimental research designs to broad conceptualisations like the following:

> "…evaluation is the determination of the merit, worth or significance of some phenomenon (policy, programme, process, system, organisation, delivery arrangement, output, outcomes, set of relationships, etc").[1]

In terms of policy advice, evaluation can focus on several questions, including:

- Was it a good piece of advice? The evaluation would question the quality of the advice itself. This could be measured in terms of what constitutes good quality policy advice including:
  - process issues, based on the notion that the right process will lead to a good product (have the correct steps and processes – including consultation, peer review, quality assurance – been carried out?)
  - analytical quality (assessed against quality standards for policy advice[2])
  - by using 'tools' such as measures of ministerial satisfaction, 'hit rate' of advice accepted, and level of influence of advice over final decision).

- Was the intervention stemming from the policy implemented as planned? This type of evaluation focuses more on the implementation process and can be done as a programme progresses. It is often referred to as formative evaluation.

- How did the policy impact on or contribute to the achievement of government objectives or overall outcomes? This evaluation would evaluate the impacts of the advice on broad outcomes. This type of evaluation is often referred to as summative evaluation. Because it is hard to show the level of causality between advice, ensuing decisions, and impacts, and because many factors affect outcomes, outcome evaluation has tended to be put in the 'too-hard basket'. Some argue that the NZPMM has done little to lift it from that position, and indeed may discourage outcome evaluation.

This paper focuses on the third type of evaluation, and the extent to which it is a feature of policy advice and ex-post review of policy settings. For the purposes of this paper, outcome evaluation is defined as any systematic attempt to measure the impact of some Government intervention, or mix of interventions, on desired outcomes.

---

[1]   Aitken, Judith, "The way we carry out the Queen's business: big issues to be dealt with over the medium term: information and evaluation", in *Future Issues in Public Management*, p.40

[2]   as set out in, *Purchase Agreement Guidelines*, The Treasury, Wellington, 1995/96, p.35. Other agencies have taken these quality guidelines and adapted them to their own purposes.

Is outcome evaluation an essential aspect of good quality policy advice?

> *"The "intelligent state" is one "...in which well developed policy advice -- both in output choice and outcomes feedback – is the central ingredient".* [3]

Evaluation is about improving the quality of decisions by informing current and future advice with lessons learned from previous experience. Evaluation provides an important learning and feedback mechanism. In terms of policy advice building evaluation criteria into policy proposals forces a focus on objectives, or outcomes, sought (ex-ante). It compels advisors to think about what they want to achieve and how they will know when and if the policy or intervention they are proposing has succeeded or not. Ex-post evaluation is an important tool for identifying policy successes and failures. It therefore should help to identify where resources are not being put to the best use and could be redirected.

All policy advice implies a view on causality. It is legitimate, therefore, to ask that that view be made explicit; for advice to identify intended outcomes, the risk that they might not be achieved, and the risk of unintended outcomes occurring instead. (intervention logic). It is also legitimate to ask that some criteria for testing this logic is included in policy advice (evaluation criteria). Ex-post, evaluation provides valuable information to be fed into future advice. Evaluation then, plays an important role in the policy process. It has the potential to provide a better understanding of the impacts of interventions and how they affect key outcomes. As such it offers a stronger basis for new policy. As the Department of Labour evaluation guidelines argue:

> *"The advice given to Ministers on policy, and the decisions made by managers on how to implement policies and programmes, can be (and frequently are) based on pure a priori reasoning or intuition. But a priori reasoning and intuition, on their own, are not adequate. Having a process for systematically testing whether a policy or plans behave as predicted is likely to help us develop better policy and plans so that, next time around, performance is improved. This, in a nutshell, is why evaluation is important".* [4]

Without evaluation, the quality of policy advice to Ministers is likely to be sub-optimal. So is evaluation happening? If not, why not?

---

[3]   Matheson, Alex, *Policy: the business of relating outputs to outcomes*, address to AIC conference on Managing Quality Policy, 7 March 1994, p.12.

[4]   *Evaluation Guidelines,* NZ Department of Labour, LMPG, 26 October 1995, p.4.

Current practice: to what extent are departments evaluating the impact of policy advice outputs on outcomes?

A good deal of evaluation activity occurs in the Public Service. Many departments engage in evaluation[5] of some kind. However most are focused on evaluation for the purpose of better delivering and implementation of programmes. Less emphasis has been placed on evaluating the impact of interventions on broader outcomes or on how departmental activities contribute to Government's stated policy priorities. Indeed, the dominant adjective used by senior policy managers to describe the extent to which departments are evaluating the impact of their policy advice on outcomes, was "modest". This message was confirmed by empirical research on a sample of Cabinet papers conducted by the State Services Commission. That research attempted to assess the extent to which policy proposals presented to Cabinet included a proposal to evaluate results, or at least included an intervention logic that was transparent enough to allow evaluation to be carried out at some later point. That research showed that only 7% of policy proposals suggested any ex-post review of results. Evaluation is typically not built into the policy advice at the outset, thereby making future review problematic.

A number of departments develop advice within an analytical framework, and this is seen as a substitute for using data and the results of evaluations. Feedback loops (data from operations), observation, and overseas experience are often regarded as substitutes for evaluation. Where evaluation does occur, it tends not to be systematically integrated into policy processes.

Senior policy managers cited a range of reasons why evaluation of outcomes is not a strong feature in the New Zealand context.

Low perceived demand

Some policy managers complained that ministers were not interested in evaluation. While they were keen to know the 'winners and losers' in any policy proposal, they were less interested in monitoring the longer term impacts. Their reluctance to support evaluation was evident in purchase agreement negotiations. A number of reasons were tendered for this reluctance. It was claimed that in some cases ministers were loathe to say precisely what they wanted to achieve, and evaluation is difficult if the outcomes sought are not clear. Others suggested that ministers just wanted to look ahead rather than backwards. This may be because they are reticent about setting their own policies up for future scrutiny, and/or because they are not interested in spending resources on scrutinising the success or failure of initiatives undertaken by predecessors. Ministers change and new ministers tend to be less interested in the results of evaluations started under a previous minister. Moreover, in particular:

> *"…social impact evaluations are not liked as they can be politically embarrassing".*

---

[5] Departments engaged in evaluation include: the Department of Internal Affairs; the Department of Social Welfare; the Education Review Office; the Ministry of Education; the Ministry of Justice; the Department of Inland Revenue; and Te Puni Kokiri.

It was claimed that evaluation gets crowded out by other, immediate demands from ministers.

Fluctuating policy settings

There appeared to be some reluctance to evaluate on the basis that when policy settings change rapidly (as is often the case in some sectors) there is little point in evaluating past or current settings.

> *"You can't have everyone evaluating what happened in the past because the world is always moving on, and new policy is needed for that new situation. Policy settings don't stay set for the long-term; we're in an environment of continuous refinement and review"*.

Funding

Funding for evaluation is an issue. Managers claimed that if evaluation was not built into original policy proposals then it was almost impossible to get funding. This reinforces the need to build evaluation in up front. But it also raises issues about the extent to which reviews of long-term established policies (as opposed to new initiatives) are inhibited. Some interviewees complained that Treasury was apt to quash any long-term evaluation proposals. One agency claimed that their proposal for a five-year evaluation was cut back to three years by Treasury, despite compelling international evidence that five years was an appropriate time-frame for results to be realised. The long-term nature of evaluation of social policy outcomes – often requiring longitudinal research – was posited as one reason for the dearth in evaluation in the social policy areas. Funding and budget issues are discussed further below.

Methodological issues

Methodological problems are also a factor in policy managers reticence towards evaluation:

> *"…evaluation never provides uncontroversial answers. All social policy evaluation is plagued by the problem of the counter-factual – you never have a control. All experience suggests it is expensive, difficult, and controversial"*.

Evaluation against outcomes is considered just too hard. Some of those interviewed felt that other tools - such as good information flows from operations - provided better information more easily, for example:

> *"..feedback loops are more important than academic evaluations"*.

Dearth of skills and capabilities

The lack of competent evaluators was cited as a reason why evaluation did not get done:

> *"There aren't many good evaluators to do the necessary work…"*.

Self interested stakeholders

Concerns about how the results of evaluation might be used were also raised.

> *"There are two ways to run a review; 1) you go as gang busters, looking for someone to blame, or 2) you take a more gentle approach, focused on learning, with no blame to be attributed".*

Criticisms about ministers' unwillingness to be scrutinised might be mirrored in the Public Service: public servants might be equally attached to certain policies and reluctant to see them challenged.

The 'good news': promising practices

Despite the difficulties and concerns described above, there is evidence to suggest that departments are starting to think more positively than in the past about outcome evaluation. Moreover, some have made considerable progress and stand as examples of promising practice. Examples include: the Labour Market Policy Group (LMPG) of the Department of Labour (see box); the environment indicators work developed as part of the Environment 2010 Strategy; the outcomes work developed in the Transport Sector; and the strategic research and evaluation issues being developed in the Department of Social Welfare.

Department of Labour: Evaluation Guidelines

The Department of Labour has a requirement that evaluation is part of all new policy development, with funding for evaluation built into baselines. All new policy proposals include an evaluation plan of which the first main section requires a description of:

- New Policy Objectives and Intended Outcomes
- Criteria for Assessing Accomplishment of Outcomes
- Evaluation Objectives
- Evaluation Methodology
- Evaluation Audience
- Monitoring Reports

- Implications for Monitoring and Operational Systems
- Constraints
- Consultation
- Timing
- Resourcing
- Budget

Moreover there is an understanding that in addition to new policies, existing functions should also be periodically evaluated. There is a three-year evaluation plan maintained by the LMPG of which an evaluation unit forms a separate unit. [6]

Despite these promising practices, most policy managers describe their evaluation work, particularly related to outcome achievement, as rudimentary or non-existent. It

---

[6]　In the evaluation literature there is a lot of debate about the relative merits of internal evaluation (to ensure buy-in and learning) and external evaluation (to avoid self-interest and capture). LMPG, in taking the middle ground (internal but separate) has perhaps found a useful and workable compromise.

is therefore useful to ask whether there are some institutional barriers or something inherent in the current public management model that inhibits evaluation against outcomes?

NZPMM: inherently hostile to outcome evaluation?

Outputs versus outcomes

As early as 1994, Alex Matheson, then a senior manager in the SSC, highlighted some of the issues arising for public servants from the dichotomy between outputs and outcomes in the NZPMM.

> "*We have until recently comforted ourselves with the curious view that departments are responsible for outputs and ministers are responsible for outcomes – and that therefore the lack of attention to outcomes – the impacts of policy on society – while deplorable, is not really our responsibility. I say curious because it slides over the public servant's vital role of advising ministers on whether activities are worth pursuing or not. Public servants do not decide outcomes – that is for ministers, but for that matter public servants do not decide outputs, ministers "buy" them. In any case the "not our responsibility" blanket is not keeping us very warm at present*".[7]

Much has been said about the relationship between outputs and outcomes[8] and the distinction made between the two in the *Public Finance Act 1989*. The Act is based on the principle that ministers are accountable for outcomes, while departments are accountable for the production of outputs. The distinction in responsibilities for outputs and outcomes allows clear lines of accountability – which can be written into contracts – between each department and their minister. Outcomes have been thought too broad for contract specification. Difficulties in measurement and causality have been given as reasons why it is impossible to hold departments accountable for outcomes, and therefore why outputs and not outcomes should be the focus for departments.

By 1992, there was a growing sense of disquiet that New Zealand public managers had become so focused on outputs, that is, what they were doing and how they were doing it, that they were starting to lose sight of the reasons why they were doing things. That is, they were losing sight of the outcomes Government was seeking. The risk in this is that we might be doing things well, but perhaps doing the wrong things. Some even argued that New Zealand was out of step with international practice

---

[7]   Matheson, Alex, *Policy: the business of relating outputs to outcomes*, address to AIC Conference on "Managing Quality Policy", March 1994.

[8]   See for example: Ball, Ian, *Outcome Specification*, Public Sector Convention 1992; Rankin, Peter *Performance and accountability in the public sector: outcome measurement*, State Services Commission 1994; Matheson, Alex, *The output/outcome nexus*, State Services Commission 1995; Matheson, Alex, and Blakeley, Roger, *Evaluation of strategic planning from the public sector perspective*, address to Australasian Evaluation Society Conference 1996; *Lifting the game: from outputs to outcomes* proceedings of the Public Service Senior Management conference, State Services Commission 1997.

because it was not trying to incorporate outcomes more formally as strategic drivers in the public management system.

The SRA/KRA strategic management system was set up in 1994 to help improve links between outputs and outcomes. The SRAs were designed, according to the then Prime Minister to "bridge the gap between the document, *Path to 2010*, and the one-year focus of departmental budgets and chief executive performance agreements".[9] There was still no formal process, however, by which ministers were informed of the impact of the outputs they were purchasing on the outcomes, for which they were intended. In a paper commissioned by the State Service Commission in 1994, Peter Rankin suggested that outcome evaluation was necessary, not simply for strategic purposes, but also to check upon the effectiveness of purchased activities:

> *"…outcome specification and measurement is an important requirement for improved State sector management at all levels from government's strategic objectives through to programme evaluation".*[10]

In 1995 a group of senior officials from the Treasury and the State Services Commission convened meetings to examine the outcome dilemma. They conceptualised the problem as:

> *"The focus on outputs while improving the accountability and transparency of government/departmental processes has not assisted the decision-making of policy makers, policy advisors and the public".*[11]

The group produced a draft paper, highlighting the lack of information about and focus on outcomes. Advice was never presented to Ministers and the group was disbanded in early 1996. The reticence about suggesting a greater focus on outcomes reflected a fear that such a focus would divert attention from outputs and could potentially undermine the foundations of the performance-accountability system. This suggests some definitional rigidity, and narrow conceptualisation (albeit easy to say in retrospect) that did little to further the ultimate goal of ensuring that outputs contribute to the achievement of outcomes. The arguments underlying this thinking deserve to be unpacked.

A renewed focus on outcomes, and in particular more evaluation of the impacts of policy outputs on outcomes, should not detract from the current outputs based performance management system. It should not be seen as a zero-sum game: that to focus on outcomes you lose sight of outputs. It is the links between the two that are important. Moreover, it does not follow necessarily that for departments to focus on outcomes they should be held accountable for their achievement. Both ex-ante identification of the impacts of outputs on outcomes and the ex-post evaluation of the same, could be seen as a complement to the current accountability system. Instead of getting tied up in debates about accountability we should be focusing on the real goal; finding mechanisms to improve the quality of policy advice to Ministers. Enhanced

---

[9]   Strategic Result Areas for the Public Sector 1994-97, 1994, p3
[10]   Rankin, Peter, op cit.
[11]   Introductory paper on outcomes, 28 May 1995.

information on the links between outputs and outcomes should therefore be a key target.

From accountability to responsibility

With the benefit of time we can also see that earlier disquiet that an outputs focus was acting as a disincentive on outcomes may also have been overly pessimistic. In fact we might have been tying ourselves up in semantics. In interviews with senior policy managers none cited the "not my responsibility" argument as a reason why they were not evaluating against outcomes, or not doing it well. Indeed, as noted above, some departments have put considerable effort into establishing the links between outputs and outcomes despite not being held accountable for the latter. Perhaps departments are less compliance-driven than central agencies have reckoned. For example, in a speech to the Public Service Senior Management conference (1998) Stuart Macdonald[12] outlined the conceptualisation of accountability underlying the outcomes work in the Transport sector:

> *"Accountability means that government agencies have a responsibility to influence outcome results. This does not mean writing into a contract or a performance agreement that you are going to be 100% accountable for reducing an accident rate to a certain level by a given time. It is a matter of recognising that there's a responsibility to influence the outcome result that's being sought".*

In the end, policy agencies are responsible for delivering policy advice. The responsibility to conduct policy reviews is part of the quality of that advice. Agencies that are not providing information on the links between outputs and outcomes are arguably not performing against the quality criteria for policy advice (including those laid out in the SSC's "The Policy Advice Initiative" booklet.[13]

Budgets and short-termism

Other systemic factors may operate more as a disincentive to a focus on outcomes than the outputs/outcomes dilemma, and may in fact have been overshadowed by it. Budget issues are not strangers to the outcome evaluation debate. At a general level, current budget arrangements skew energies towards outputs and short-term gains at the expense of medium to long term achievement of outcomes. A former Minister of State Services, Paul East,[14] warned that outcomes might be lost in the drive for expenditure control:

> *"If the budget round becomes overwhelmingly to do with expenditure control – and particularly to do with short-term expenditure control – the incentives for intelligent debate about the achievement of high-level outcomes are bound to be driven out...."*

High level debate, including on the continuation of existing policies and programmes, can only be based on information about what works and what doesn't – information

---

[12] Macdonald Stuart, *Manage your outcomes*, PSSM conference proceedings, State Services Commission 1997, p.50.
[13] The Policy Advice Initiative: opportunities for management, State Services Commission, 1995.
[14] Hon Paul East, Opening Address, PSSM conference proceedings, State Services Commission, 1997, p.7.

that is lacking in the absence of evaluation of policy against outcomes. Unless we look back we might be simply treating symptoms, not causes. New policy could be introduced to deal with what might be the unintended impacts of previous policies. So what are the incentives to evaluate in current budget arrangements?

In theory, there are in-built incentives in current budget arrangements to reprioritise activities. With a multi-year budget focus and known parameters on spending (e.g. an expenditure cap imposed through a new administration's 'coalition agreement') departments have relative certainty over expected resources. With limited potential for additional spending, any new initiatives usually have to be financed through reprioritisation. This would suggest a strong incentive to evaluate existing programmes to determine where reprioritisation might occur. As Bushnell[15] noted:

> *"An effective budget constraint is a wonderful spur to seek better advice"*.

However, the picture painted by policy managers portrays a different story. Intended incentives do not seem to be working. The perceived incentives are to hold on to what you've got, to carry on business as usual while shouldering any reductions in resources, and to avoid open scrutiny of the relative merits of existing programmes. In short, policy managers perceive few incentives to evaluate.

Contrary to theory, current budget arrangements may indeed create weak incentives to reprioritise and therefore to use evaluation as a tool for indicating where reprioritisation might occur. The budget process focuses on scrutiny of new initiatives, rather than existing outputs, which are not usually questioned unless a specific review is commissioned. There might be pressure to absorb the costs of some new initiatives, but this accounts for a fraction of overall expenditure.

The incentives to use evaluation to reprioritise between votes are even less evident, especially given the implied risk of losing your department's programme in favour of someone else's found to be more effective. The same lack of incentives applies to the review of cross-cutting policies, that involve outputs from different agencies aimed at the same outcome.

The short-termism inherent in the system is also not conducive to outcome evaluation. Evaluations – especially in the social policy area – typically require a longer time-frame or some ongoing commitment to monitoring progress over time. There is little provision for this long-term commitment (more than three years) in current budget arrangements. And, as noted above, at a very specific level, policy managers cited resource issues as a factor inhibiting evaluation against outcomes. They felt they simply couldn't accommodate the costs of evaluation in current baselines and knew that the possibilities of obtaining additional funding were slim.

Therefore, while the current public management system may not inhibit outcome evaluation, it is not overly conducive to it either. Departments perceive the incentives to invest resources in creating information on the impacts of interventions to be weak.

---

[15]   Bushnell, Peter *Does evaluation of policies matter?* paper presented at a workshop on Policy Evaluation, Foreign and Commonwealth Economic Advisors, London, February 1998, p.10.

If it is agreed that outcome evaluation is an essential component of quality policy advice then we need to examine the necessary conditions for the development of an evaluation culture, and where there might be greater incentives for increasing evaluation activity.

Conditions to encourage more evaluation: possible incentives

The following attempts to define some necessary conditions and to suggest mechanisms for encouraging and enabling departments to develop a capacity for outcome evaluation.

Demand conditions

- Ministers should demand ex-ante information on how the policy outputs and resulting interventions they are purchasing will contribute to the achievement of outcomes, and ex-post information on progress in working towards those outcomes. In doing so they should be mindful of timing issues; that results may not be realised in the short term and that evaluation requires medium to long-term commitment and investment in information systems.

- Incentives to encourage ministers to focus on reprioritisation between low and high value outputs and therefore to demand more policy review should be investigated. Options might include mechanisms for holding ministers more accountable for outcomes. Currently Ministers are required to submit outcome statements as part of the Estimates, but there is no corresponding ex-post reporting requirement.

- Even in the absence of demand, public servants should supply ministers with evaluative information as an integral part of policy advice. The obligation to provide quality, 'free and frank' advice (and notions such as the duty of care) deems it necessary. Public servants need to be conscious of the disincentives on ministers to call for such information, and should not wait to be asked for it. However there is a potential tension here between the provision of evaluative information and situations where a Minister has specifically declined to purchase such information.

- Central agencies should use levers in their realm of responsibilities to encourage evaluation and policy review, and to signal the importance of outcome information (such as the State Services Commissioner's *Letter of Expectations* to chief executives including questions about the extent to which they have evaluation mechanisms in place). This provides a signal as to the importance of evaluation; an important first step, given the mixed, sometimes negative, messages central agencies have given to agencies in the various debates over outcome evaluation. This signal should be mirrored in similar letters of expectations originating in the Department of the Prime Minister and Cabinet (in SRA/KRA consultations) and the Treasury. These would need to be backed-up by complementary ex-post

review mechanisms.   As the agencies responsible for cross-government
advice to ministers, they might also from time to time propose strategic
reviews of existing policy settings, and how they are contributing to
outcomes (and to determine whether others might contribute more or
better).

- Budget processes and Cabinet Office rules might be amended to encourage
  evaluation in proposals related to significant new policy. To ensure that
  evaluation is cost-effective, and to avoid overkill, some guidelines
  (including what constitutes "significant") might be required in this area (e.g.
  fiscal materiality, importance of policy in relation to key government
  objectives (SRAs), new or controversial type of intervention etc. – this
  notion is expanded below). In developing these mechanisms it would be
  important to be vigilant about compliance costs (which would suggest an
  other then mandatory approach). In any case, if agencies decided not to
  build evaluation proposals into policy advice they should propose some
  alternative mechanism for checking effectiveness.

Stable policy settings

Ministers must be aware that when policy settings change constantly it is virtually
impossible to evaluate the impacts of certain interventions. A volatile policy
environment undermines the potential to learn from past experience. It also sends
signals to public servants that evaluating what they are doing is pointless. If policies
are likely to change regardless of information about whether current polices are
working or not, there is no incentive to generate such information.

But, public servants must remember that policy change is a political reality. Where
the change is minor, or tinkering around the edges, this should not provide an excuse
to abandon evaluation. Rather it poses a methodological challenge to be addressed. It
is also an argument for using a mixture of evaluation techniques so that if it becomes
impossible to evaluate actual outcomes, information about how a type of intervention
works in practice can still be generated.

Enhanced outcome specification

Ministers need to be more explicit about the outcomes they are seeking. It is hard to
track progress towards a goal when you don't know where the 'goal posts' are. As
was argued in a recent report on strategic social policy:

> *"The Government's statements of strategic intent – the SRAs - while a major
> step forward in terms of highlighting high level areas of priority, are not
> sufficiently explicit as a basis for strategic policy objectives, and do not
> readily translate into operational policy objectives".* [16]

---

[16]   State Services Commission, *Strategic social policy*, report to the Minister of State Services, 2
         March 1998.

However, while it is up to ministers to determine which outcomes government wishes to pursue, this does not let departments 'off the hook' in terms of refining those broad high-level outcomes down to the level of workable and measurable policy goals.

Departments therefore need to become more skilled at outcome specification, in consultation with ministers, in order to dis-aggregate broad outcomes into a hierarchy of lower level outcomes. In this way they can develop information systems for tracking progress in meeting those objectives. They also need to develop better systems for working with other agencies which also have a role in meeting the same or related outcomes (e.g. through shared KRAs) where there might be synergies to exploit, or policy conflicts to avoid or minimise. The *Environment 2010 Strategy* is a good example of this type of process. Building evaluation into policy proposals has the spin-off benefit of forcing ministers – in signing-off policy proposals – to also agree explicitly to the outcomes they wish to achieve.

Information and indicators

Enhanced sources of outcome information need to be developed. While the above demand factors might encourage better information gathering within departments, no one agency has the full picture of the links between social policy areas issues. There is also a need for some cross-government information generation. This would include the development of 'outcome indicators', especially in the social policy area (e.g. indicators of social cohesion) where we are currently the least well endowed. At present, some time-series analysis is limited by a lack of historical data in some areas. Previous attempts to create organisations to produce outcome relevant information, particularly in social policy (such as the demise of the Crown Research Institute dedicated to social policy) have failed. The results of the applied social science research deliberations should go some way to improving this situation. Evaluation activities would both draw on this information and feed into it.

Robust evaluation methodologies/frameworks

Robust evaluation methodologies and/or frameworks need to be developed, especially given the attributional problems inherent in attempting to show the impacts of policy advice on eventual outcomes (both in ex-ante policy advice and ex-post reviews). Current good practice in the environment sector and Labour could be disseminated for the benefit of departments that are less well developed in the area.

As noted above, one of the reasons for the reticence associated with evaluation against outcomes is the extent to which it is difficult to demonstrate causality between policy advice, ensuing policy decisions and later outcomes. A multiplicity of factors impact on the shape of final outcomes, many of which are outside the control of government. Other methodological issues make evaluation problematic. For example, experimental or quasi-experimental research conditions are difficult to create in social policy areas; the use of control groups can raise ethical issues.[17]

---

[17] Ethical issues are raised in a number of areas related to evaluation, including the process of conducting an evaluation. See *Guidelines for the Ethical Conduct of Evaluations*, Australasian Evaluation Society Inc., 1998.

The choice of method depends on the type of intervention to be evaluated, the variables to be included, and the complexity of environmental factors. Evaluators need to be aware of the limitations of the various approaches, as do policy managers. Good communication between policy managers and evaluators is imperative to ensure that appropriate evaluation criteria is built in to policy proposals up front.

Outcome evaluation should emerge from the 'too hard' basket. Just because evaluation is hard, and methods are not perfect, does not mean it should not be attempted. There is some difference of opinion on the level of certainty required in results. Bushnell concludes *that "If the effects of a policy cannot be identified then the evaluation should not be carried out, since ...the evaluation will add no value".*[18]

Certainty about causality is almost always impossible to determine. But decision makers might not need 100% certainty. If the correlations between an activity and the desired outcome are high, and if plausible alternative explanations can be ruled out, then there is a reasonable basis to argue that an intervention was effective.

Several policy managers admitted that their attempts at evaluation against outcomes was rudimentary, but that the information it provided was better than none at all. In the final analysis, building evaluation into policy advice forces advisors to think about the causal linkages and impacts on outcomes, and it enables future reviews of those policies and their underlying assumptions to be tested, ex-post. As Christopher Pollitt has explained:

> *"In practice the conduct of such [ex-post] research is greatly assisted if the programme concerned is possessed of a clear statement of its intervention logic. That is, do the foundational documents explain in some detail exactly how the actions taken by public agencies are supposed to produce the desired outcomes? If the intervention logic is clear it becomes much easier for evaluators to establish whether the causal linkages out there in the real world closely match those envisaged in the intervention logic, diverge from it somewhat, or whether the two bear little resemblance to each other... Outcome evaluators also need to remain aware of the sometimes large influence of implementation contexts. It is not just the purity of the internal logic of the policy that leads to the desired outcomes, but the interaction between the programme logic and local contexts... Thus evaluators have to develop methods of enquiry which characterise contexts as well as interventions".*[19]

---

[18]   Bushnell, *op. cit.*, p.4.
[19]   Pollitt, C., "Looking outcomes in the face; the limits of Government action", in *Lifting the Game: from outputs to outcomes*, proceedings of the Public Service Senior Management conference, State Services Commission, 1997, p.14.

Skills and capability

There is a need to develop capability in a range of evaluation and review techniques, including a skills base in departments, in order to either carry out evaluations internally or to manage external evaluations appropriately.

The lack of evaluation capability was noted by policy managers and has been highlighted by commentators. There appears to be insufficient capability within the Public Service itself, while some concerns were also expressed about the quality of external evaluations. Judith Aitken - who as Chief Executive of the Education Review Office, the only public organisation dedicated to evaluation, has borne the brunt of this skills shortage more than most - has characterised the state of affairs as:

- *"Our evaluation skills are not only fairly amateur, they are generally lacking in professional rigour...*
- *"We can locate very few training opportunities for those interested in evaluation...*
- *"We can find almost no academic courses in evaluation...*
- *"In the Public Service we lack evaluation-specific skills, techniques, institutions, professional codes and, most of all, professional practitioners".*[20]

If public managers see, and are given, signals that evaluation is important, they will begin to invest in this capability. This highlights the important links between the various conditions necessary for developing an evaluation culture.

However, evaluation is not a panacea

The above discussion is not intended to suggest that evaluation is a panacea, or that everything should be evaluated, or that evaluation is the only tool for generating information for ministers on whether government activities are contributing to the achievement of outcomes. So when should evaluation be used? Should we evaluate everything?

Australian experience – which could be characterised as "evaluating everything that moves" – paints a picture of evaluation overkill (interestingly, the Australians have recently moved away from this blanket approach).[21] Evaluation costs money.

---

[20]   Aitken, Judith, *Future Issues*, p.44.

[21]   The Australians have moved away from mandatory evaluation of every programme every 3-5 years to ad hoc strategic reviews and a requirement that all new policy proposals include evaluation criteria. Their previous system seemed to lack focus and tended to encourage a compliance mentality (resulting in more but not necessarily higher quality evaluation), and was very agency based. There was little, if any, evaluation of policy settings or mixes of interventions. The Australians are now working on mechanisms for building incentives to evaluation into chief executive reporting arrangements. On the upside, previous mandatory evaluation has created an evaluation culture. The Australians are facing the same methodological issues as NZ, especially in terms of developing an outcomes focus, and have identified similar skills and capability gaps. For an opinion on the evaluation strategy and its impact on resource allocation, see Michael Di Francesco, "The measure of policy? Evaluating the evaluation strategy as an instrument for budgetary control", *Australian Journal of Public Administration*, vol.57, March 1998.

Longitudinal studies and controlled experiments are especially expensive. Evaluation might be the fall-back option once less expensive alternative mechanisms for testing the merits of activities have been considered.

For small projects results might be based on good cost/benefit analysis and overseas evidence, rather than a controlled evaluation. Given the size of the New Zealand population it might be legitimate to use evaluative information generated in other jurisdictions: in others, unique local conditions might necessitate purpose built research. Sometimes existing statistics might be indicative. A lot of information (e.g. demographic, mortality data) is collected for other purposes, but might be used as indicators of progress in moving towards certain outcomes, thereby lessening the need for a purpose-built evaluation of policy settings. But, certain questions need to be asked: What would the gap be in information in the absence of evaluation, and is it worth investing in evaluation?

Evaluating everything is not cost-effective, especially small, or business-as-usual activities. Some guidelines as to when to evaluate would be useful. The State Services Commission could work with the Treasury to develop some framework to guide agencies. When to evaluate might depend on a range of factors. For example, the fiscal dimension (is a lot of money being spent?), whether it is a new or potentially controversial type of intervention (level of risk), and the extent to which the policy relates to the key priorities of government. Resources for evaluation should be targeted at strategic areas. The role of central agencies – as suggested above – might even extend to identifying where reviews of existing policy settings - mixes of interventions that are outside the domain of a single organisation - might be in order. Departments themselves will need to rise to the challenge of building evaluation into new policy proposals.

The extent to which the results of evaluations are fed into the policy process is also an important consideration. If you are not going to act on the results of evaluations then there is little point in doing it. But, even if the results of evaluations do not lead to major change in policy settings – for political or other reasons – they still provide useful learning information for policy advisors on the impacts of various interventions and hence contribute to the foundations of learning organisations, and a learning government.

Conclusions

Previous discussions around outcome evaluation mirrored related debates about the extent to which a greater focus on outcomes might undermine the current outputs-based performance management system. These fears reflect what might be seen as a systemic obsession with notions of accountability; that departments should not be expected to do anything for which they were not to be held to account.

If we put accountability to one side and focus instead on the notion that public servants should be expected to show that what they produce actually contributes positively to the strategic goals a government is trying to achieve; and that information on the impacts of interventions is crucial for future decision making, then the arguments for more evaluation become less problematic and more compelling

(methodological and measurement issues notwithstanding). Evaluation should be encouraged as part of good quality policy advice that is essential for ensuring that decision making is based on critical information about the impacts of interventions on desired outcomes.

General References:

- Uhr, John, and Mackay, Keith (eds), (1996) *Evaluating policy advice*, Australian National University, Canberra

- Chapman, Ralph, (1995) "Improving policy advice processes: strategic outcome definition, policy evaluation and a competitive policy market", *Public Sector*, Vol.18, No.2, pp16-21.