



**DEPARTMENT
of the PRIME MINISTER
and CABINET**



Te Puni Kōkiri

STATE SERVICES COMMISSION
Te Komihana o Ngā Tai Kāwanatanga



THE TREASURY
Kaitiaki Take Kōwhiri

Learning from Evaluative Activity

Enhancing Performance through Outcome-focussed Management

Prepared by the Steering Group for the Managing for Outcomes Roll-out 2004/05
November 2003

CONTENTS

Overview	3
Introduction.....	4
Spreading Evaluative Capability Across the Organisation.....	5
Evaluative Strategy, Plans and Prioritisation	6
Choosing the Best Approach	8
Putting It Together in a Learning Organisation	9
Appendix 1 - Critical Success Factors for Organisations Building Evaluative Capability	11
Appendix 2: Using the 'Results Chain' for evaluation and performance monitoring	12
Appendix 3: Further Reading	13

Overview

Managing for Outcomes (MfO) is about enhancing focus and performance through a more strategic and outcome-focussed approach to management and reporting. As part of seeking continuous improvement, agencies are expected to learn from success and failure and make the necessary modifications to their output mix and delivery.

This paper provides guidance on how, when and where to use evaluative products to improve management decisions. It emphasises planning and prioritising evaluative activity, monitoring results, and building evaluative thinking into management practice.

Evaluative activity is essential to good management and strategic thinking. It includes all forms of learning activity that aim to improve focus, delivery and impact. All parts of your agency – senior executives, managers, planners, operators and analysts - must work together to gather data, process results, build evidence bases, and use new knowledge to improve decisions and interventions if you are to obtain the most benefit from evaluative activity.

All departments should have evaluative strategies laying out a medium-term work plan, information needs (incl. administrative data), and how results will drive decisions. Departments engaged in joint activities are expected to develop joint evaluation plans.

Evaluative strategies should focus evaluative activity on ‘big’ areas where there is uncertainty about what works and evaluative work can generate useful information. They focus effort on the strategic and operational priorities of an agency, its sector and the Government. They anticipate, and aim to inform ‘big’ decisions. They aim to provide assurance that strategies and interventions are delivered well, and produce desired results.

Outcome-focussed organisations learn from their experiences, even when this involves bad news. Openness to unexpected results – and willingness to change – will ultimately determine how well New Zealand’s public service actually serves the public.

Most evaluative effort should focus on major interventions where audit, monitoring and evaluation is cost-effective — and where reducing uncertainty about delivery and results can benefit New Zealanders. You do not need to assess every intervention. But it should be clear why you focus where you do, and why ‘big’ areas are omitted.

Agencies are expected to report planned evaluative activities in Statements of Intent, and report major findings from evaluative activities in annual reports and on websites. Staff should be avid users of results. Results should influence or change what you do.

You should actively seek Ministerial buy in to your evaluation strategy as part of your ongoing Ministerial engagement. Keep Ministers informed about progress, risks and significant results.

Introduction

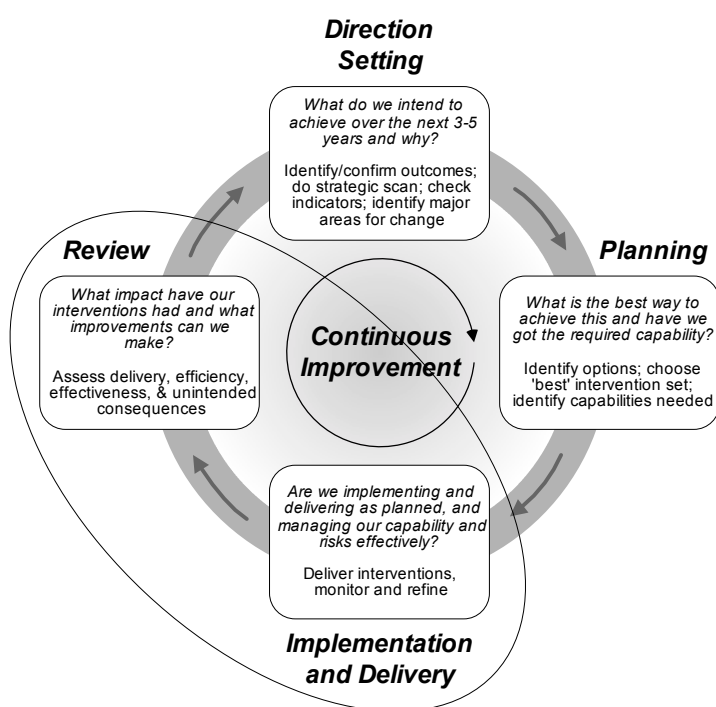
1. This paper provides guidance to management on how, when and where to use evaluative thinking to improve decisions on strategy and outputs. Governments act to improve life for their country’s citizens and residents. Democratic governments must be transparent about what they want to achieve, and how they know their actions achieve results. The aim of the Managing for Outcomes (MfO) initiative is to create a management culture that is fact-based, results-oriented and open, where agencies are accountable for:

- helping Ministers decide which outcomes to pursue, and their relative priority;
- advising Ministers on which interventions to purchase, based on value and cost;
- specifying interventions, and delivering interventions to agreed standards;
- managing major risks to achieving outcomes, including using evaluative activity to increase certainty about how well interventions are working;
- adjusting strategies and outputs on the basis of evaluative findings.

2. A lot of what is discussed in this paper is simply good practice presented to us by departments. For some, the concepts will represent an ideal that has yet to be reached. But agencies must build on the best of what we are already doing, alone and together.

3. The management cycle of an outcome-focussed organisation encompasses direction setting, planning, delivery and reviewing results. Results are then fed back into the cycle to inform future planning (Exhibit 1). The cycle drives ongoing improvement of interventions and the output mix.

Exhibit 1: The Good Management Cycle



4. Early guidance stressed direction setting and planning. New guidance extends the focus to delivery, review, and using information to improve direction, plans and outputs¹.

5. Ongoing improvement of strategies and programmes requires knowledge of where issues and opportunities lie, what can improve outcomes, and what cannot. Better results will come from better design, delivery and review, and using performance feedback to improve decisions and actions.

¹ Managing for Outcomes Guidance 2003, <http://www.ssc.govt.nz/display/document.asp?NavID=208&DocID=3516>

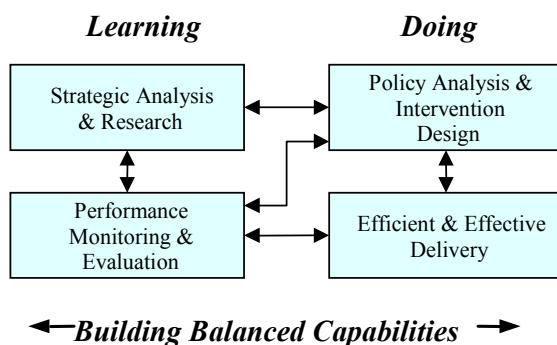
6. Thinking about how, when and where to get evidence of what works (and why) and what does not must happen around the cycle and across all areas of management. This should not be delegated lightly. To boost performance, all parts of your agency – senior executives, managers, planners, operators and analysts - must build evidence, process results, and use the knowledge gained to drive decisions.

Spreading Evaluative Capability Across the Organisation

7. Evaluative thinking drives the systematic improvement of strategies, interventions and business practice. It starts with acknowledgement that problems may exist, occurs through the systematic analysis of strategies, interventions and business processes, and ends with reports tailored to the needs of decision makers across your agency.

8. An interesting question posed to us was whether the issue is limited evaluative capacity, or whether solutions lay in the habits, mindsets and attitudes of planning, policy and evaluation professionals. Answers may vary across agencies. Irrespective of your answer, evaluative work is inextricably linked to good management practices.

Exhibit 2: Creating a Culture of Enquiry



9. Evaluation is only one bit of the evaluative jigsaw (Exhibit 2). Evaluative work includes self and peer review by planners, intervention designers and operators. It is embedded in applied research, monitoring and audit work.

10. Growing evaluative capacity means thinking about how managers manage, and how resources are concentrated, deployed and used. Challenges will not be met just by adding more people, funding, or business units.

Contrary to the common wisdom that measurement is the most enduring problem in Managing for Outcomes, communication and utilisation of information in decisions are equally problematic. Problems are most evident in government-wide efforts. There are relatively few cases in which [agency] systems provide wide dissemination and integration for decision making purposes.²

11. A recent review³ of the state sector found that the use of evaluative thinking to inform strategy, policy, service delivery and budget decisions was patchy. The review suggested there is a need to build capability in five ways.

- *Creating a 'culture of enquiry'* — creating demand from ministers and senior managers for 'quality' evaluative activity, producing good and relevant analyses, and getting leaders to use the findings to inform decisions.

² After 'Paths to performance in state & local government' p.214-5. Maxwell School of Citizenship & Public Affairs, USA.

³ Doing the right things and doing them right: improving evaluative activity in the state sector. Joint Treasury & SSC report, September 2003. <http://www.ssc.govt.nz/display/document.asp?docid=3506>

- *Educating policy and programme managers* — ensuring people involved in planning, policy design and delivery know when and how to promote evaluative activity, how to commission and manage it, and how to use results.
 - *Prioritising evaluative effort* — focusing effort on major strategies and outputs.
 - *Building evaluative capacity & skills* — conduct high quality evaluative activity.
 - *Coordinating agencies' collective efforts* — sharing data, methods and findings.
12. Staff must know about the principles and practices underpinning effective intervention — and know enough about evaluative practices to be ‘informed purchasers and consumers’. Agencies with a culture of enquiry:
- Nurture capabilities needed to recognise success and failure
 - Protect what works, while seeking and evaluating new ways of doing business
 - Value analytical rigour and evidence above rhetoric and corporate norms
13. Above all else, a strong culture of enquiry is driven by:
- Decision makers that seek, spread and back robust evidence
 - Clarity of purpose, action and intended results
 - Ability to recognise success in qualitative and quantitative terms
 - Willingness to question the utility of everything you do
14. A key challenge is to shift how evaluative findings are viewed from threat to opportunity. We need forward looking, prospective discussions of how and where we can do even better into the future.
15. A culture of enquiry is useful only if is backed by a management culture that promotes performance information and responds appropriately to unexpected news. MfO is about changing culture and focus. Evaluative thinking is a very useful tool.

Evaluative Strategy, Plans and Prioritisation

16. All departments should develop strategies and plans for their evaluative activities. Strategies should focus on gathering information to make ‘big’ decisions in areas where evaluative techniques can provide useful information. Departments engaged in joint activities are expected to develop joint plans. Plans should cover existing activities as well as new outputs and new initiatives. This requires effort and careful planning.

17. It is often better to answer a few ‘big’ questions well, than to spread effort thinly. The goal is not simply to look at where the ‘big bucks’ are spent. Not everything needs to be evaluated. As a generalisation, however, departments should initially focus effort on strategies, interventions and activities that either:

- May have a big impact at modest cost (i.e. warrant protection or expansion);
- May pose risks to an agency or its clients and stakeholders;
- Consume significant resources, either in total or per intervention; or
- Will, if pilots work, expand to contribute (or cost) a lot.

18. Evaluative activity is valuable only if you are *willing* to respond to bad news. Assuming this basic condition is met, prioritisation systems account for the:

- Significance of the strategy or intervention in fiscal or outcome terms;
- Level of uncertainty about design, delivery and results;
- Traction expected from methods and information available; and
- Likely cost of the proposed evaluative activities

19. Prioritisation is typically subjective not precise, and is iterative in nature. As you firm up detailed plans for individual interventions (below), new information on the cost and likely usefulness of evaluative findings often make you revise your earlier priorities. Some intriguing and important questions simply cannot be answered at a useful time.

20. Steering Group expects you to ask five main questions about major strategies and interventions:

- Are they designed well, aligned to real needs and likely to improve outcomes?
- Do the quantity, quality and coverage of outputs match the intervention logic?
- Are near-term results and outcomes improving in the areas anticipated?
- How strong is the evidence that change resulted from the strategy or outputs?
- Are there any unplanned consequences (positive or negative)?

The foundations of good evaluative work are the foundations of good management:

- *Clear outcomes, rationale and policy ('good policy frameworks')*
- *Intervention logic applied to both strategies & interventions during the design phase*
- *Capability to identify the right questions, analytical approach, and time to evaluate*
- *Information gathered from the start, to support operators, managers & evaluative work*
- *Generation of evaluative results in the timeframes needed to support decisions*
- *Tight communication of results & recommendations, & use in decision making*

21. An early step in evaluating a particular strategy or intervention is to identify questions to be addressed (above), when they can be answered, and good methods for gathering evidence. Then decide whether your questions can be answered with reasonable precision and cost, and within a useful timeframe. This involves:⁴

- Familiarisation with the strategy/intervention and a review of relevant literature
- Rationale analysis (intervention logic, incl. why government must intervene)
- Critical review of available data and information, and how you can fill gaps
- Selection and timing of evaluative actions by analysing options and costs

The results of this work should be used to develop evaluative strategies and plans and decide where your department will put its evaluative effort first. Results should also be used to clarify what desires of management cannot be met and manage expectations.

⁴ See Figure 1: The Flow of Evaluation Tasks, *Programme Evaluation Methods, Measurement and Attribution of Program Results*, Third Edition, Treasury Board of Canada Secretariat, March 1998.

22. Plans for individual studies should clarify what will be evaluated or monitored, how and when, over a multi-year time horizon if necessary. Intervention logic can help you firm up plans and information needs. Logic models help you clarify what can be assessed, prioritise and sequence work, and limit costs of evaluative activity.

23. Not everything lends itself to measurement. Evaluative specialists must review management's 'wants' and give advice on which questions can be answered now, the methods likely to be cost-effective, and which questions cannot be answered well – at least for now. In some cases new data and information is needed, and plans must be put together to ensure affordable information becomes available into the future. But start by thinking through how existing data, such as administrative data, could be used.

24. Evaluative strategies should often focus first on major interventions where evaluation is cost-effective — and where risks or uncertainty surround their outcomes. But over time, build information about how well other strategies and outputs work.

25. The Steering Group recognises the challenges of evaluative activity and good decision-making. It is not demanding conclusive proof of impact (but hopes to be surprised on occasion). It is looking for confidence that your 'big' interventions work. The goal is to protect or grow interventions that work, improve design, fix problems in a timely manner, acknowledge uncertain results, and focus effort on productive activity.

26. In the process be alert for unintended outcomes not covered by your monitoring. Tunnel vision can occur around intended results, leaving blind spots⁵.

Choosing the Best Approach

27. Robust intervention – not evaluative action – is the key to success. Not everything needs a full evaluation. Evaluative options include monitoring, self-review and performance audit as well as formal evaluations of policies and programmes.

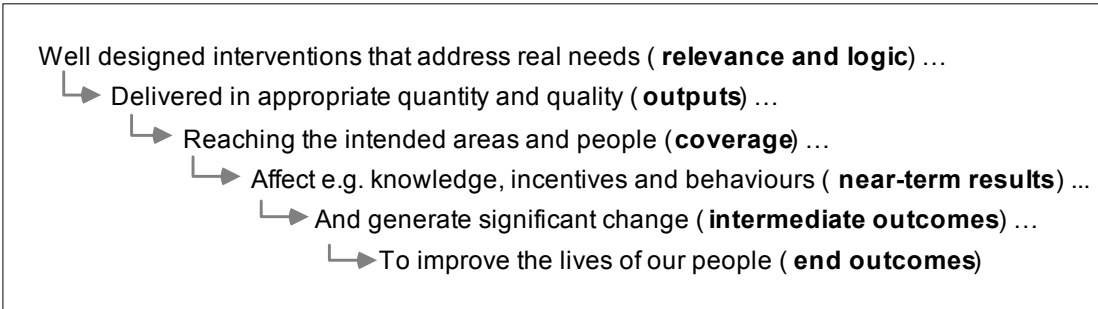
28. Conclusive proof of impact is rare. But the harder it is to show impact, the more critical it is to show that your intervention rationale is robust, your design is good, and that your interventions (outputs) are delivered in line with your rationale and design.

29. Intervention logic provides one framework — the 'results chain' — for identifying and monitoring how interventions work. Intervention logic can be used to identify some results that you could assess to build confidence that outputs 'work as planned'. Feedback can drive improved delivery — often well before end outcomes are known.

30. The concept of the 'results chain' is straightforward. Behind every action is a set of assumptions about how your actions (outputs) improve outcomes. Assumptions can be tested if you organise information and reporting systems properly.

Exhibit 3: The “Results Chain”

⁵ One US study praised completion rates in mediocre schools, because it overlooked poor outcomes for expelled pupils.



31. By monitoring indicators along the ‘results chain’ you can get early information on how well and where interventions work, and where change is needed. Monitoring outputs, coverage and near term results helps you take early action if problems are found. Monitoring gives you basic information on outputs and impact that you need to assess efficiency and effectiveness. And when impact has been assessed, monitoring results provide a mix of measures that show where and why things worked or failed.

Putting It Together in a Learning Organisation

32. Agencies need to be transparent about *what* results they are trying to achieve and *how* they know specific activities or programs achieve results. The goal of evaluative activity is to improve delivery and certainty things work. At very least, agencies should know if their interventions cause harm.

33. Policy design, implementation, ongoing delivery and management must focus on what results are expected, and how evidence of performance will be gathered and brought together for decision making. Learning organisations modify their strategies and outputs in response to what they learn from success and failure.

34. Managers must ensure evaluative products inform crucial decisions on direction and intervention. Managers and staff across the organisation must act as informed purchasers and consumers of evidence, including research and evaluative findings.

35. The first step in building a ‘culture of enquiry’ is putting what you do to the test. You need a culture that accepts short-term risk, to reduce long-term performance risks. Even partial gains in knowledge about how and why programmes work or fail help us make better decisions about if and how programmes should be delivered in the future.

36. Good decisions are driven by good information. Good information focuses decision makers minds on key issues affecting performance, and starts the search for remedies if change is indicated. It is technically robust, and credible to management. Always look at how representative results are. Do not confuse anecdote with evidence.

37. The goal is to improve decisions by reducing uncertainty about direction and interventions. This goal is most likely to be achieved by careful planning, intelligent use of data (including administrative data), and focussed use of your analytical capabilities. Different findings can be produced in different time horizons. So you need to revisit historical decisions as new results become available, and knowledge changes.

38. To sustain investment in evaluative activity, decision makers and stakeholders - including ministers – must see good information (even on bad results) drive decisions. Spend time building evaluative outputs into your decision making processes and cycle. Consider how you will ensure results are seen as objective, accurate and impartial⁶.

39. Above all else, learning organisations confront issues and change for the better. Staff at all levels must create demand for evidence showing them how to perform even better in priority areas, and be willing to act on results, both positive and negative.

40. Good results are easy to handle. We just continue doing what we did before. Openness to unexpected and poor results – and willingness to change – will ultimately determine how well we serve the citizens and residents of this proud country.

⁶ E.g. establish clear protocols and rules, ensure integrity (vs. independence), set standards and improve data. These steps should be documented for your own use, and in case the Auditor-General wants to review your results.

APPENDIX 1.

Critical Success Factors for Organisations Building Evaluative Capability

41. The Ministry of Economic Development recently worked with the Department of Labour and Ministry of Social Development to identify interdependent factors that must be nurtured for evaluative results to start shaping an organisation's work.

42. **Senior management must commit to:**

- Building capability;
- Providing ongoing, active encouragement and support to build the right climate for evaluation (including symbolic actions and consistent signalling); and
- Leading internal demand for well-targeted evaluative results.

43. **Evaluation support staff** perform multiple roles, including:

- Developing an evaluation strategy. Work with the senior management group and senior policy staff to develop prioritisation guidelines and decide what to evaluate;
- Design and advise on ways to carry out evaluation. Develop systems, methods, frameworks, tools, guidelines and data management protocols;
- Help managers develop evaluation plans, and coach them in the skills needed to develop and execute plans and use results to best effect;
- Provide technical expertise to support managers run internal evaluations, manage external evaluation contracts, and manage data collection; and
- Provide independent QA on evaluations at the request of branches.

44. **New skills** are often needed:

- Evaluators have specialist skills that need to be built and spread (quantitative and qualitative skills, and knowledge of research and evaluation methods); and
- Managers need assistance from specialists to design effective evaluation plans and build their own capability to manage internal and external evaluative work.

45. **Systems** must be implemented that provide incentives to encourage evaluation and support to execute it well. E.g. guidance for managers, a phased approach to developing evaluation plans for major policies and programmes, and systems for identifying and collecting the data needed for evaluation⁷, on a case-by-case basis.

46. **Build a culture that sustains and improves your capability over time.** This often involves building analytical and evaluative capacity iteratively over several years. Avoid creating compliance pressures and unrealistic expectations that can produce disillusionment or a backlash against the evidence-informed culture you are creating.

⁷ This includes looking at the utility of existing information before new information is gathered. Other organisations suggest management information systems should not just contain the data essential to the agency's transactions. They should also house data needed for evaluation, such as descriptive, classification and intervention data.

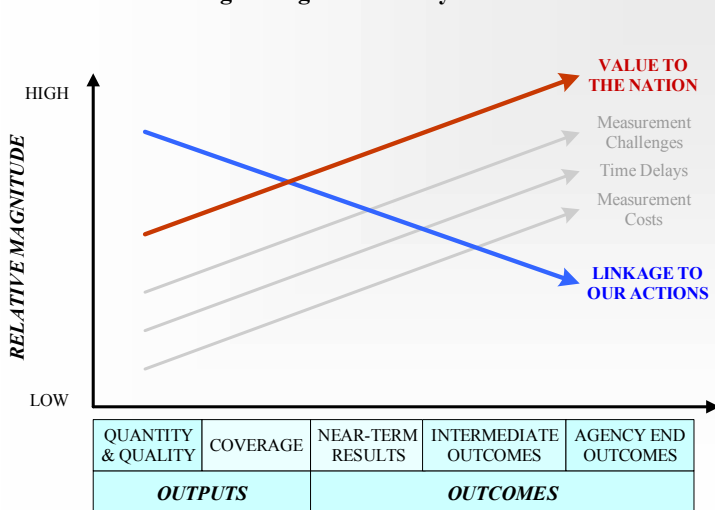
Appendix 2

Using the ‘Results Chain’ for evaluation and performance monitoring

47. Conclusive proof of impact is rare. The harder it is to show impact, the more critical it is to show the right outputs are delivered to the right groups, and the interim results suggest your intervention rationale is working. Intervention logic provides one framework — the ‘results chain’ — for identifying and monitoring these key results.

48. Performance monitoring based on intervention logic starts from a simple premise. Outcomes arise because interventions trigger a chain of events – the ‘results chain’ (Exhibit 3). Broken links often flag problems in the logic or delivery that must be fixed before interventions

Exhibit 4. Balancing Strengths: Delivery vs. Outcome Measures



are likely to improve outcomes. By monitoring results along the ‘results chain’ you can build feedback to drive robust management decisions.

49. Base decisions on the best information you have. Information on outputs, coverage and near-term results can often be gathered quickly and simply. Assessing outcomes is demanding, and often takes more time and

effort (Exhibit 4). Both forms of information are invaluable.

50. Output information describes what we control. But outcomes are what we want.

51. Over time, build more and more information on the likely effectiveness of your major strategies and interventions:

- a) Improve rationale, design and delivery using methods like intervention logic, and explore options by studying successful interventions from around the world.
- b) Confirm interim results such as output delivery, coverage and near-term results.
- c) Where possible, confirm impact by assessing intermediate and end outcomes.

a. *Check Intervention Rationale and Design*

52. Even before an intervention is made, evaluative approaches such as intervention logic allow you to improve the rationale and design, and identify measures of success.

53. Every strategy and intervention reflects a theory of cause, action and effect (Exhibit 3). These theories or logic models require testing, frequent revisiting and redesigning. This means acknowledging that ideas must sometimes be discarded.

54. Both the theory and structure of interventions should be examined. Check that the outcomes are still relevant to your direction. Be vigilant for more cost-effective ways to achieve the same outcomes. Reviewing the logic model, experience and the literature may help staff identify areas (big and small) in which delivery or coverage can be improved, or groups (such as Māori) for which the intervention is less suitable.

55. Logic models include the measures of success later evaluative activity can focus on. The models thus help you decide what could be measured, what can be measured, and what data must be collected to assess results. Use this information now to design for future evaluations and later heartache. The information you need to assess success can often be collected most cheaply during delivery. More importantly, logic models help staff fix weak aspects of the design before delivery even starts. Formative evaluation often focuses on logic models.

56. Where outcomes are difficult to assess, it is particularly important to confirm that interventions are well designed, and that outputs get delivered well. This is true of regulation and programmes. Logic models identify the critical things you must do to get results. The models show what managers must monitor, manage and strive to improve.

b. Confirm Interim Results (Output Delivery, Coverage and Near-Term Results)

57. Results matter. Behind every action is a chain of assumptions about how your actions (outputs) improve outcomes. Assumptions can often be tested if you organise information and reporting systems properly. The first step in checking assumptions is simply to see whether major results arose as predicted.

58. Feedback on end outcomes can be hard to obtain. But intervention logic can be used to identify other results that you can assess to build confidence that outputs ‘work as planned’ — often well before end outcomes are known.

59. By monitoring indicators along the ‘results chain’ you can get early information on how well and where interventions are working, and where change is needed. Monitoring outputs, coverage and near term results helps you take early action if problems are found.

60. The success of good strategies and interventions depend on good delivery. Poor outputs have poor outcomes. So you need to improve output specification and monitoring, and confirm the quantity and quality of outputs. Having monitored outputs for years, agencies must ensure outputs have the right coverage and results. If you do not get the planned coverage and near term results, your own intervention logic has been violated. Output and coverage issues are under management’s direct control.

61. Even if outcomes can be assessed in the near-term, information on outputs, coverage and near-term results provide early warning of problems, specific information

on where problems lie, and helps focus remedial action. If feedback on outcomes is slow or elusive, information on these interim results can show you are ‘on track’. When end outcomes cannot be assessed, managers must show that interim results match what advocates and intervention designers said must happen for outcomes to improve.

c. *Wherever Possible, Confirm Outcomes*

62. Outputs still matter. But outcomes matter more. There is widespread agreement that failure to improve outcomes indicates a need for change. The challenge is to assess outcomes (impact) accurately, in time to inform critical decisions.

63. Measuring impact can be very challenging. Impact assessments are often ambiguous. Methods for measuring and attributing changes in outcomes (intended or not) to strategies and interventions are beyond the scope of this work. But attribution remains one of the most challenging issues in assessing impacts.

64. Even if attribution is strong, take care in generalising findings. Impact should be reported with sufficient information on context to support good decisions. The accuracy and relevance of findings depends on when, how and where information was gathered.

65. Interventions that clearly improve outcomes can often be protected from major change. But by fine-tuning design, delivery and coverage results can often be improved further. This may involve improved targeting, or ‘tweaking’ services to improve acceptability or effectiveness. If results are well below expectation, outputs should be modified or reduced until results improve. Interventions with clear adverse impacts must be stopped or redesigned.

66. Interventions that produce modest but positive results cause a dilemma. You need to decide whether performance can be improved sufficiently. But if you continue, you may want to constrain growth until performance meets expectations. Interventions with conflicting results also pose a dilemma. Variation between groups may signal a need for tighter allocation, but variation may also indicate monitoring problems and signal a need to make decisions cautiously.

APPENDIX 3.

FURTHER READING

The Steering Group recognises the value of different types of evaluative activity, and the need for flexibility about when different methods should be applied. No single evaluative method can meet all the needs identified in this paper. The following papers provide useful information on, and discussion of, different evaluative approaches.

Evaluation Methods:

Programme Evaluation Methods, Measurement and Attribution of Program Results, Third Edition, Treasury Board of Canada, Secretariat, March 1998.

http://www.tbs-sct.gc.ca/eval/pubs/meth/pem-mep_e.asp

The Magenta Book: Guidance Notes for Policy Evaluation and Analysis, Prime Minister's Strategy Unit, Cabinet Office, July 2003.

<http://policyhub.emps.gov.uk/servlet/Menu?id=1724>

Results Chain Monitoring:

John Mayne, *Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly*, Office of the Auditor General of Canada, June 1999.

http://www.oag-bvg.gc.ca/domino/other.nsf/html/99dp1_e.html

Modifying the Output Mix

Pathfinder, *Building Block 5: Maximising Outcomes from Our Interventions*, July 2003.

<http://io.ssc.govt.nz/pathfinder/information.asp>