

evaluation – monitoring

Monitoring

The slow cousin of evaluation, or an equal partner?

Evaluators often regard monitoring as playing a secondary and relatively simple role compared to evaluation proper. This paper argues that this view underestimates the potential of monitoring information in enhancing the value of evaluative work, in particular to increase the 'half-life' of evaluation findings. Moreover, it suggests the possibility of a more dynamic interchange between monitoring and evaluation. Specifically, monitoring complements the fragmented and ad-hoc nature of evaluation work, so that the process of monitoring presents an opportunity to develop a framework within which individual evaluations can exist. Drawing on program logic and CMO theory, the key to this interchange is to build of systematic body of knowledge and theory that drives, and is in turn informed by, evaluation and monitoring. The work within the Department of Work and Income New Zealand provides an example of an attempt at implementing such a framework.

Marc de Boer



Introduction

Over the past 25 years, two approaches have emerged to provide empirical information to decision makers on the effectiveness of social programs. One is evaluation, which uses principles of social science research to assess the concept, design and implementation of programs (Rossi & Freeman 1994). The other is monitoring¹, which provides decision-makers with timely information on a program's progress, often against set goals or benchmarks.

Several authors within the discipline of evaluation have raised concerns over the apparent independence of these two approaches and how they often come into conflict within the decision-making process (Bernstein 1999; Blalock 1999). The challenge these authors pose is how to marry the different information approaches of evaluation and monitoring. This paper shows one attempt, focussing particularly on the contribution that monitoring can make in the generation of organisational knowledge.

Marc de Boer is Senior Research and Evaluation Analyst with the Centre for Operational Research and Evaluation in the NZ Ministry of Social Development.

Criticisms of evaluation and monitoring

Before outlining how monitoring and evaluation might work together, it is useful to first examine the weaknesses of each.

Among evaluators at least, evaluation is regarded as the best means to judge and understand the impact of programs on outcomes. But this strength is also an Achilles heel, in that good evaluations need expertise, resources and, above all, time. This often leads to a lagged cycle of commissioning evaluations to address policy questions, only to have evaluations reported well after the necessary decisions have been made. Conversely, decision-makers often view earlier evaluations as out of date; and, rightly or wrongly, irrelevant to current policy questions.

This game of catch-up also produces an incoherent body of work. The limited timeframes of individual evaluations often preclude review of previous evaluation or research findings². By acting in isolation, such evaluations have limited opportunity to contribute new insights into the policy or program under review (Lipsey 2000; Anderson 1998). This failure to accumulate evaluative knowledge is ironic, as this is one of the cornerstones of scientific inquiry upon which evaluation bases its legitimacy.

In contrast, the perception of monitoring information, especially among evaluators themselves, is that it is second-best to evaluation proper (Blalock 1999; Davies 1999). Monitoring information is overwhelmingly quantitative and is usually of what can be measured easily, often leaving important aspects under-represented³. Further, monitoring analysis often comprises only simple descriptions of program operation and outcomes, reflecting its primary role for public accountability and the focus on the 'what' rather than 'why' questions (Newcomer 1997). Evaluators see a particular risk in the use of monitoring information in assessing program impact, believing that monitoring can fail to properly address, or even acknowledge, issues of causality; potentially misleading decision-makers as to the effectiveness of programs (Mayne 1999; Blalock 1999)⁴.

The positive conclusion of this rather negative introduction is that the weaknesses and, by implication, the strengths of evaluation and monitoring are complementary (Perrin 1998). Synthesis of evaluation and monitoring serves to increase the value of both in the decision making process (Davies 1999). The question is how this can be best achieved.

Role of theory

It is argued in this paper that grounded theory is the key to linking evaluation and monitoring. Moreover, such theory operates at two levels, the first being at

the individual program level, while the second places programs within the broader setting of the problem or issue being addressed.

At the program level, good design will outline

In contrast, the perception of monitoring information, especially among evaluators themselves, is that it is second-best to evaluation proper (Blalock 1999; Davies 1999). Monitoring information is overwhelmingly quantitative and is usually of what can be measured easily, often leaving important aspects under-represented.

how a program is to be implemented (inputs, processes and outputs) as well as clearly state the logic and assumptions about how the program's outputs will influence people's outcomes in the desired way. It is also important to understand the context within which programs and policies operate (e.g. other policies/programs, resource availability) and to deal with the competing theories of how programs influence outcomes (Wholey 1977; Pawson & Tilley 1999; Funnell 1997; Scheirer 2000).

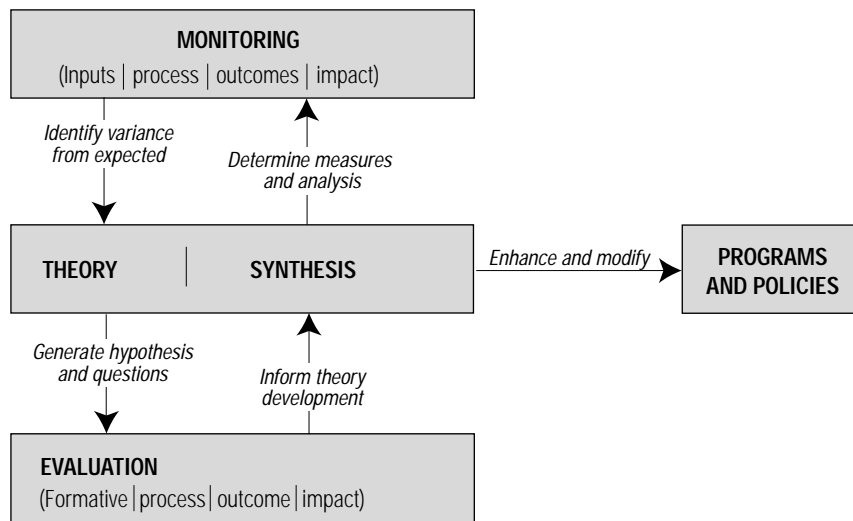
The second level of theory positions the intervention logic of programs within mid-level theories (Pawson & Tilley 1999) of the social or economic problem being addressed. This provides an opportunity to be able to trace how the outputs of specific programs influence broader outcomes, and to contrast this with alternative programs and initiatives. Such issue-focussed theory helps to anchor program evaluation and monitoring within a broader body of research; enabling analysts to use this framework to better understand the processes that contribute to the observed impact of individual programs.

Accordingly, monitoring and evaluation play central roles in testing and expanding program theory. Put simply, monitoring describes 'what is', while evaluation seeks to understand 'why this is so'. The ideal would be that after construction of a program's intervention logic, a monitoring framework would be implemented to measure critical aspects of a program's intervention logic. Sufficient care needs to be

taken to ensure the framework is credible to external stakeholders and that decision makers understand the validity, reliability and limitations of the measures used (Blalock, 1999)⁵. Where monitoring signals the program is not working as intended or not having its expected impact, or where competing causal theories cannot be ruled out, then evaluation can be applied (Perrin 1998, 1999). Such evaluation will either confirm existing theory or produce new understanding, which in turn can redirect monitoring effort (Perrin 1998). Therefore, a cycle

Put simply, monitoring describes 'what is', while evaluation seeks to understand 'why this is so'.

FIGURE 1: SCHEMATIC OF THE PROPOSED RELATIONSHIP BETWEEN EVALUATION AND MONITORING



emerges of continuing monitoring with periodic evaluation (Mayne 1999), each enhancing the robustness of the program(s) theory (Figure 1).

Such a monitoring/theory/evaluation cycle increases the value of both monitoring and evaluation information for decision-making. Firstly, it allows evaluators to respond better to the information needs of policy-makers, lessening the need for ad-hoc or piecemeal evaluative research. In particular, it can help assure control agencies of the effectiveness of policies or programs by using monitoring information for accountability as well as program improvement (Mayne 1999; Scheirer 2000; Bernstein 1999). Secondly, it enables identification of those areas where organisational understanding is limited, so that scarce evaluation resources are better directed.

Evaluation by this logic is necessarily intensive and targeted, while monitoring needs to be broad in coverage, and occur in an ongoing and timely manner. This does suggest that methods of monitoring are less sophisticated and rigorous than for evaluation. However, this does not imply that monitoring information must be limited to simple counts of inputs, processes, outputs and outcomes (Scheirer 2000). Rather, monitoring information needs to be sufficient to provide stakeholders with assurance that the program is working as intended, and provide credible evidence of the effect it has on outcomes (Mayne 1999). The key aim of monitoring, therefore, is to minimise the resources needed to provide credible and reliable information to inform decision-makers of the operation of programs. This equation will be a function of the policy/program being evaluated, available information, and the technical expertise of the

analysts involved.

The following section (pages 11 and 12) provides an example of a first attempt at setting up this framework and the role that monitoring information plays within it. However, while theoretical frameworks strive for simplicity and elegance, practice is always a messier affair. Nevertheless, the example that follows will hopefully illustrate three points:

- possibility for the use of sophisticated methods as part of program monitoring;
- the importance of theory in understanding the information produced;
- the iterative nature of building organisational knowledge through evaluation and monitoring.

Conclusions

The work done so far within DWI represents the first iteration of our application of the monitoring/theory/evaluation cycle. How successful it will be in completing the cycle remains to be seen. However, what I hope the paper will show is the value of developing sophisticated monitoring techniques in combination with theoretical frameworks to help support and direct evaluation effort. To this end, monitoring and evaluation become equal partners in developing and synthesising knowledge at all levels of the organisation.

Notes

- 1 The literature discusses monitoring under a number of headings, most often as performance measurement and management information systems. The term used in this paper primarily refers to any regular source of

Monitoring of employment programs

The Department of Work and Income (DWI) is responsible for administering income support and providing of employment assistance, and was created through the merger of Income Support Service and New Zealand Employment Service in 1998⁶. This merger provided an opportunity for the integration of administrative data on income support and employment information. This, coupled with the rapid advancement in microprocessing power and analytical applications like SAS, has allowed internal evaluators to have direct access to administrative data for evaluation and monitoring.

Outcomes and impact of employment programs

One of the key questions for government is whether the employment assistance DWI provides is effective in helping disadvantaged job seekers. As a result, DWI was directed to review the effectiveness of nine of its largest programs. The internal evaluation team saw this as an opportunity to implement a monitoring framework focussed on the outcomes and impact of employment programs provided through the Department.

The two most significant challenges in setting up this monitoring framework were finding a reliable outcome measure and estimating the counterfactual. In both cases, the initial outcome measure and the estimation

technique were challenged by external agencies. Because the monitoring framework had to be credible to external agencies, DWI took considerable care to work through the issues raised. This work, and its acceptance, has improved both the utility and robustness of the analysis.

In the end a simple 'Independence of DWI' measure was the outcome measure chosen, which reflected whether a job seeker was receiving either income support or employment assistance from the Department. The counterfactual was estimated using a quasi-experimental design. This involved constructing propensity-weighted comparison groups, producing a group of non-participants who have the same characteristic profile as that of the participants. The similarities between the two are strongest for those characteristics that distinguish participants from the average job-seeker population (de Boer 2001b). The top-line results of the analysis are summarised in Table 1 below (de Boer 2001a, 2001c).

It is not the purpose of this paper to go into detail over the results of the findings. However, I should point out that they illustrate the risk of interpreting gross outcomes as a proxy for the program's impact. This point is perhaps best illustrated by the example of Job Connection (wage subsidy targeted at the long-term unemployed), which has low outcomes but the highest impact ratios.

TABLE 1: PROPORTION OF PARTICIPANTS INDEPENDENT OF DWI AND IMPACT RATIO BY PROGRAM AT 12 AND 24 MONTHS AFTER PARTICIPATION START

Program type	Program name	Participants independent of DWI ¹		Adjusted impact ratio ²	
		12 months	24 months	12 months	24 months
Wage subsidy	Job Plus	68%	70%	1.64	1.29
	Job Connection	41%	43%	1.99	1.46
On-the-job training	Job Plus Training	52%	58%	1.27	1.16
Work experience	CTF/Community Work	26%	41%	0.83	0.94
	Task Force Green	47%	56%	1.31	1.18
	Job Plus Maori Assets	65%	61%	2.05	1.33
Self-employment assistance	Enterprise Allowance	76%	70%	1.92	1.26

¹ Independence of DWI is where a job seeker is no longer receiving a core benefit or participating in employment programs.

² Impact ratio: estimated using propensity weighted regression and is the ratio between the proportion of participants and non-participants Independent of DWI, controlling for other observable jobseeker characteristics.

Base: Includes all program participants who started between 1 January 1998 and 1 July 2000.

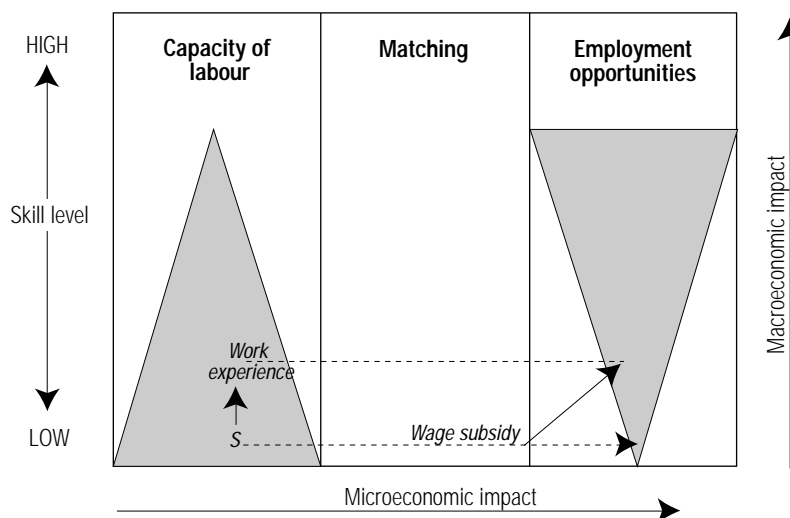
employment programs

Theoretical model of the labour market

Developing consistent measures of program outcomes and impact immediately focuses attention on why differences exist. The first high-level attempt at explaining this is done by placing these programs within a simple model of the labour market (Figure 2 below). Figure 2 divides the labour market into three parts, 1) Capacity of Labour, 2) Matching, and 3) Employment Opportunities, and assumes that unemployment is largely structural. That is, there is a mismatch between the supply and the demand of labour at each skill level (represented by the two triangles). Different types of programs can be placed within the model according to how they are intended to address unemployment within the labour market (as indicated by the solid arrows).

Although simple, the model is a powerful way of illustrating how employment programs attempt to address unemployment. More importantly, it draws the links between the observed microeconomic impact and their potential macroeconomic affects⁷. For example, wage subsidies are located close to the boundary between matching and employment opportunities, and therefore are expected to result in high outcomes for participants and thus have a significant microeconomic impact. On the other hand, work experience programs seek to develop the capacity of labour, and therefore the link to opportunities is more tenuous. Accordingly, the microeconomic outcomes and impact of these programs would be more modest. Both these conclusions are borne out by the monitoring data presented above.

FIGURE 2: CONCEPTUAL FRAMEWORK OF THE LABOUR MARKET WITH STRUCTURAL UNEMPLOYMENT



However, the model points out that assisting participants into employment is only part of the picture. What the monitoring does not provide is information on the quality of the outcomes achieved (represented as an increase in the skills of jobseekers and the types of work they move into). Whether programs assist job seekers into higher skilled employment has implications for the ability of programs to address structural unemployment, as well as for the associated risk of displacing or substituting people who are equally disadvantaged in the labour market as those assisted.

The model suggests that there is a close alignment between the observed microeconomic impact of work experience programs and their

potential macroeconomic impact. That is, if the program fails to improve the capacity of labour, then it is unlikely that there will be a positive impact, microeconomic or macroeconomic. However, this relationship does not hold for wage subsidies, where a program may still achieve a high microeconomic impact but, because of poorly targeted employers, may not result in increased capacity of labour (as indicated by the dotted arrow in Figure 2). In other words, the program simply cycles disadvantaged job seekers into and out of low-skilled employment. Based on this theoretical model, the monitoring of microeconomic impact for wage subsidy programs (as presented in this paper) is insufficient on its own to assure decision-makers that such programs are operating effectively.

What has been discussed so far is the first iteration of the monitoring/theory cycle (see Figure 1). From this, several new evaluation and monitoring initiatives have been proposed. The analysis of macroeconomic risks signals the need to monitor more closely individual employers' use of wage subsidy programs, in order to reduce the risk of participants substituting or displacing other disadvantaged job seekers and workers. Moreover, this work also identified a large gap in the Department's knowledge about the role that wage subsidies play in persuading employers to hire disadvantaged jobseekers. This places the research and evaluation team in a strong position in recommending that subsequent evaluation effort should be directed into these areas.

labour market

- information or data on programs rather than the management systems that they support.
- 2 To be fair, this is often compounded by the difficulty of getting evaluations into the public domain, particularly internal evaluations critical of an organisation's programs or policies.
 - 3 As Perrin (1998) points out, 'many activities in the public policy realm, by their very nature, are complex and intangible and cannot be reduced to a numerical figure ... What is measured or even measurable, often bears little resemblance to what is relevant'.
 - 4 This does not even begin to address the more serious issues with monitoring when linked to unrealistic performance targets producing goal displacement and other perverse behaviours, which can also seriously undermine the veracity of the information gained through these systems. In this respect, Maoist China and Stalinist Russia are perhaps some of the most instructive examples of the misuse of performance management.
 - 5 Perrin (1998) makes the useful distinction between 'bad' and 'dirty' data. Dirty data is common to all research and evaluation methods. However, bad data is a specific risk with monitoring information generated for performance management purposes, where there is a very real risk that results or methods have been 'engineered' to reach set targets.
 - 6 As of October 2001, DWI was itself merged with the Ministry of Social Policy to form the Ministry of Social Development.
 - 7 Microeconomic impact refers in the main to the effect that programs have on the job seekers who are directly assisted by the program, while the macroeconomic level considers how programs may affect the outcomes of non-participants as well as participants.

References

- Anderson, Diane 1998, *What Have We Learned and Where to Next? A Review of Evaluations of Employment Programs, 1994-1997*, Labour Market Policy Group, Department of Labour, Wellington, New Zealand.
- Bernstein, David, J. 1999, 'Comments on Perrin's "Effective use and misuse of performance management"', *American Journal of Evaluation*, vol. 20, no. 1, pp. 85-93.
- Blalock, Ann Bonar 1999, 'Evaluation research and the performance management movement', *Evaluation*, vol. 5, no. 2, pp. 245-258.
- Davies, Ian C. 1999, 'Evaluation and performance management in government', *Evaluation*, vol. 5, no. 2, pp. 150-159.
- de Boer, Marc 2001a, *Review of the Subsidised Work Appropriation*, Department of Work and Income, Wellington, New Zealand.
- de Boer, Marc 2001b, *Methodology of the Review of the Subsidised Work Appropriation*, Department of Work and Income, Wellington, New Zealand.
- de Boer, Marc 2001c, *Six-monthly Review of DWI Employment Programs*, Department of Work and Income, August 2001, Wellington, New Zealand.
- Funnell, Sue 1997, 'Program logic: An adaptable tool for designing and evaluating programs', *Evaluation News & Comment*, vol. 6, no. 1, July, pp. 5-17.
- Lipsey, Mark 2000, 'Meta-analysis and the learning curve in evaluation practice', *American Journal of Evaluation*, vol. 21, no. 2, pp. 207-212.
- Newcomer, Kathryn, E. 1997, 'Using performance measurement to improve programs', in *Using Performance Measurement to Improve Public and Non-profit Programs*, ed. Kathryn E. Newcomer, New Directions for Evaluation, no. 75, Jossey-Bass, San Francisco.
- Pawson, Ray & Tilley, Nick 1997, *Realistic Evaluation*, Sage Publications, London.
- Perrin, Burt 1998, 'Effective use and misuse of performance measurement', *American Journal of Evaluation*, vol. 21, no. 2.
- Perrin, Burt 1999, 'Performance measurement: Does the reality match the rhetoric?', *American Journal of Evaluation*, vol. 20, no. 1, pp. 101-111.
- Rossi, P. & Freeman, H. 1994, *Evaluation: A Systematic Approach*, Sage Publications, Thousand Oaks, California.
- Scheirer, Mary Ann 2000, 'Getting more "bang" for your performance "buck"', *American Journal of Evaluation*, vol. 21, no. 2, pp. 139-149.
- Wholey, J. S. 1997, 'Evaluability assessment', in *Evaluation Research Methods: A Basic Guide*, ed. L. Rutmna, Sage Publications, Thousand Oaks, California.



Insurance for AES Members



IC Frith & Associates can help you with your insurance requirements

- Exclusive Professional Indemnity Facility for Evaluation Professionals
- Workers Compensation Insurance
- Business Insurance
- Travel Insurance
- All other business insurance tailored to your needs

For an obligation free quote contact:

Richard Parker or Brenda Hill
 Phone: (02) 9899 2133; Fax: (02) 9634 2396
 Email: richardp@icfrith.com.au

IC Frith & Associates Pty Ltd

ABN 48 097 088 287

Insurance Brokers