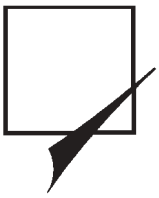
Evaluation

[1356–3890 (200301)9:1; 5–25; 032583]
Vol 9(1): 5–25



Teaching People to Fish? Building the Evaluation Capability of Public Sector Organizations

BRON MCDONALD
Natural Resources and Environment, Australia

PATRICIA ROGERS
Royal Melbourne Institute of Technology, Australia

BRUCE KEFFORD
Natural Resources and Environment, Australia

In response to an increasing demand for public sector accountability, many government agencies have sought to develop their internal evaluation capabilities. Often these efforts have focused on increasing the capacity to supply credible evaluations, yet addressing demand is just as important. This article focuses on a government agency and tracks its five-year journey towards developing such a capability. It documents contextual matters, drivers for change, the actions taken by the agency, and its response to emergent challenges during four phases. Based on feedback from project staff and managers and those involved in the capability development project, it offers seven recommendations. These are: start small and grow evaluation; address both supply and demand; work top-down and bottom-up simultaneously; use a theory of change behaviour; develop a common evaluation framework, including a generic programme theory; build knowledge of what works within the agency's context; and systematically and visibly evaluate each stage.

KEYWORDS: building capability; evaluation capability; evaluation training; evaluation use; programme theory

Introduction

Continuing interest in improving public sector performance and accountability has led to various efforts to build the evaluation capability of public sector agencies or regions. Rather than simply increasing the budget for external evaluations, these efforts usually focus on making changes within the agency. This

often means developing staff skills in data collection and analysis, but occasionally involves broader organizational change. Sometimes these efforts are motivated by a perception that there are not enough external evaluators to conduct the evaluations required (Leeuw et al., 1999) or that there are insufficient funds.

More commonly, however, efforts to build evaluation capability are justified in terms of the well-known epigram 'give someone a fish and they eat for a day; teach them to fish and they eat for a lifetime' (for example, Fetterman (1996: 11) when introducing empowerment evaluation). This encapsulates the belief that an evaluation capability will provide enduring organizational benefits, including a sustainable resource for producing evaluations as well as a system for encouraging and using evaluation.

However, this does not mean that evaluation capability should only be understood to refer to self-evaluation by programme staff and managers. Such a view increases the risk of producing evaluations that serve the interests of programme staff rather than programme clients, funders or other legitimate stakeholders (Scriven, 1997). An evaluation capability should also include effective use of external evaluations and ongoing monitoring. The aim is not just to conduct evaluations but also to commission, manage and use them. In addition to having skills and knowledge about evaluation, create an environment in which these can be deployed. The formal processes, frameworks and resources developed to support evaluative activity should be complemented by tacit knowledge, human capital, symbolic actions and learning about learning.

To return to the fishing epigram, as well as individuals skilled in catching fish, we need the equipment to successfully fish, an effective distribution system, people who want to eat fish, and an entire fishing system that is sustainable. In short, we need to focus on working with the whole organization, not just on developing the skills of individuals.

Aspects of Evaluation Capability

Supply and demand are two broad aspects of evaluation capability. Many efforts at building evaluation capability have focused primarily or even exclusively on supply – on documenting and developing the skills, tools and resources that are available to produce evaluations. For example, in their briefing notes for the Presidential Strand of the 2000 American Evaluation Association conference, Milstein and Cotton (2000) defined evaluation capacity quite narrowly as 'the ability to conduct an effective evaluation (i.e., one that meets accepted standards of the discipline).'

In a similar way, Russon and Patel's (1999) review of evaluation capacity in Africa assessed capacity solely in terms of the technical expertise of existing external research and evaluation agencies. The danger in this is the risk of ending up 'all dressed up with nowhere to go' (Williams, 2001) – capable of producing evaluations but unable to use them, or even worse, producing evaluations that are treated as irrelevant. In fact, Mackay, working in evaluation capacity development in the World Bank, has argued that 'supply is not as crucial as demand' (World Bank, 1994). As Williams (2001) has pointed out:

> all the skills, knowledge, technical expertise and experience in the world won't help . . . if the . . . program, community, organization or environment cannot sustain and nurture those skills, and abilities. (Williams, 2001)

It is therefore important to also understand and manage the demand end of evaluation capability. One of the decisions to be made in managing evaluation demand is whether evaluation should be mandatory. Making evaluation mandatory could promote a culture of token compliance, but voluntary adoption is much slower to take effect. The majority view was clear in a panel discussion on 'mainstreaming evaluation' at the 2001 meeting of the American Evaluation Association, where many speakers warned against mandating evaluation and there was no opposing view from the floor.

Davies (1999) has categorically stated that the:

> most important lesson for me is that performance management (and evaluative enterprise in general) cannot be forced on people. Attempting to impose it will likely lead to goal displacement, unreliable information and an increase in the risk that programme relevance will be diminished rather than augmented. (Davies, 1999: 157)

In addition to consideration of supply and demand, Davidson's (2001) discussion of evaluation capability is particularly useful. As well as formal policies, systems and practices, she argues that we need to consider the effect of informal practices, symbolic actions, evaluative beliefs, values and attitudes. Symbolic actions include the fate of risk-takers and tough critics, and whether 'near misses on high goals [are] rewarded more than easily clearing easy targets'.

Building evaluation capability requires attending to all of these factors. It is also clear that while external evaluators might appropriately conduct particular evaluations, and be a resource, the building of an organization's evaluation capability must be managed from within the organization.

The Context for this Case Study

This article outlines five years of work in building the evaluation capability of one particular organization. We write from three complementary perspectives: as the leader of the evaluation development project: as the chief executive officer of the organization; and as an academic who worked on the project as an external resource. This contrasts with other studies of building evaluation capability which have been written by external consultants (for example, Preskill and Torres, 1999a). Our analysis also draws on formal reviews conducted at each stage in the process.

The organization discussed in this article has many features in common with other public sector organizations: responsibility for planning and managing large, complex projects implemented across a geographically dispersed area in an atmosphere of competition for government funding and increasingly complex demands for accountability in terms of economic, social and environmental outcomes. This increased accountability is typically articulated by three questions from the general public: 'What did you do with the money?', 'Was it spent efficiently?' and 'What difference did it make?'. Performance measurement and

evaluation separately and together are important tools available to help managers answer these questions (Blalock, 1999; Davies, 1999). Apart from these common features, five factors appear particularly important in understanding the particular context of this case study.

The first of these factors was the nature of the programmes delivered and how these influenced the strategies used to build evaluation capability. This case study was based in the Agriculture Division of the Victorian Department of Natural Resources and Environment, Australia. The Division is responsible for providing services to the agricultural sector, including fundamental and applied research projects, and extension projects that work with farmers to encourage appropriate adoption of new and improved technologies. Since 1996, the Division has separated its strategic planning and project investment decision making (the investors) from project development and delivery (the providers).

The organization has a strong science research culture, with a large number of staff being trained or experienced in approaches to research, many with PhDs in science. The strategies used to build an evaluation capability drew from the Division's experiences in delivering research and extension services. Like research programmes, the process of building evaluation capability involved formally reviewing the work of others, trialling techniques, and inviting input from external expert reviewers and visiting researchers at critical stages. Like extension programmes, the building of evaluation capability involved developing materials and processes to communicate information about available technologies, with an underlying message that adoption was likely to require adaptation to local conditions and generation of new knowledge.

The second significant factor was a shift in accountability, requiring state government agencies to plan, deliver and report in terms of the triple bottom line – economic, social and environmental outcomes. The Division had previously been required to demonstrate its contribution to increased agricultural exports. Clearly, additional evaluation frameworks and techniques were needed.

The third factor was the Division's move from 2500 small projects to 60 large multi-million dollar projects. Stakeholders and co-investors became engaged in the project development process in a more strategic way, examining industry directions rather than looking at piece-meal project applications. With more staff and increased resources, it became possible for each project to start implementing its own credible communication, risk-management and evaluation strategies.

Larger projects, however, also created a higher degree of complexity, particularly in terms of the range of stakeholders associated with each project; each having a unique interest in demonstrating the value of services to their own constituencies. Consequently, the evaluation requirements of projects became more complex and more likely to change over time.

The fourth contextual factor was identified as the biggest evaluation challenge – to effectively identify and communicate the impact of projects. Four eminent Australian scientists appointed as chief scientists to the Division had reviewed the agency's research efforts and concluded that while the science capability of the Division was credible and the projects it worked on worthwhile, the potential impact of this effort had not been fully achieved. One of the four, Sir Gustav

Nossal, even went so far as to describe the science of the Division as Victoria's 'best kept secret' (Department of Natural Resources and Environment, 2001). It was hoped that evaluation could provide useful information to help overcome the blockages.

The fifth factor was personal. People who were involved in the early stages of building evaluation capability in one project moved into positions where they could influence the whole Division, becoming evaluation champions. Patton (1997: 44) has referred to 'the personal factor' in the success of individual evaluations – the presence of a champion for the evaluation. We believe the same factor applies to the building of capability.

The next section of this article describes strategies that seem to have been largely successful in this agency. Could these strategies be translated to other agencies? Or does each agency need to develop an approach to building evaluation capability that resonates with the culture of its own particular programmes and professions? The following account attempts to provide sufficient information for others to identify where circumstances may be sufficiently similar for these factors to apply.

Stages in Developing an Evaluation Capability

Looking back over the past five years, we can identify four distinct phases, although these were not planned at the beginning. Each new phase began with reflection and replanning to identify priorities, strengths, weaknesses, options, risk-management strategies and next steps. By committing to a regime of 'trial–reflect–grow', the Division was acknowledging that the task was difficult and needed to be grounded in its own organizational context. It also needed to demonstrate an approach committed to learning and to minimize the risk of wholesale staff disenfranchizement and cynicism. The first three phases were: working with one project; working with a range of different volunteer projects; and rolling out mandatory evaluation across all new initiatives of the Division. The fourth and current phase is about consolidating within the Division and responding to new challenges such as an increased demand for evaluation research and requests for support coming from other divisions and external agencies.

The following table summarizes the key characteristics of these phases and shows the growth and change in scale and strategies.

Phase One: Addressing a Specific Need (18 months)

The impetus for this phase was the need to produce a better understanding of the impact of one particular project, Target 10, a dairy extension project that provided services to help farmers increase productivity. The project had undergone some forms of evaluation but these had not answered the questions posed by stakeholders who had direct influence over the future of the project.

The effectiveness of the extension package was originally evaluated using a comparison group process (O'Brien and Hepworth, 1994). This established that,

Table 1. Building an Evaluation Capability

	Phase 1: Addressing a specific need	*Phase 2: Experimenting with volunteer projects*	*Phase 3: Mandatory evaluation for all new projects*	*Phase 4: Expansion and consolidation*
Number of projects involved	1	10	All new projects	All new projects
Time	18 months	18 months	Two years	Now and ongoing
Staff	One person	Two people plus consultants	Two people plus consultants	Four permanent positions plus two temporary (12 months) positions
Role of the evaluation support team	Evaluator; researcher	Mentor; informal trainer; researcher	Mentor; formal trainer; researcher; facilitator; special events manager	As for phase 3 but with a stronger evaluation-research role
Evaluation of work on building evaluation capacity	Independent blind review of reports	External evaluation using semi- interviews	Internal evaluation; mixed methods structured	Planned external review; continued internal evaluation
Research and development	Designed frameworks; experimented with realistic evaluation	Supported a PhD (Dart, 2000)	Undertook research on: a framework for impact evaluation; social impact assessment; and values inquiry	Strategy for expanding evaluation research being developed

on average, farmers could increase their pasture consumption by 15 percent (which loosely translated to AU$10,000 per annum at that time) if they adopted the technologies promoted in the package.

Over its first three years of operation, the project had been subjected to numerous cost–benefit analyses, and had a comprehensive database of inputs and completed activities. Funders, however, found this information did not enable them to fully understand the impact of the project. For example, cost–benefit analysis showed that AU$62m had been returned to the Victorian dairy industry as a result of the AU$5.8m invested over three years. The industry chair of the Target 10 State Executive responded to this information by asking 'if Target 10 has returned AU$62m to the industry, I want to know who's got it, and where is it?'

The need for a different sort of evaluation was clear. It seemed that without it, programmes such as Target 10 wouldn't survive.

Acting on advice from a consultant (Wissemann, 1996), it was decided to apply a programme-theory approach using Bennett's Hierarchy, (Bennett, 1977; Bennett and Rockwell, 2000). This hierarchy describes a cause-and-effect chain through the identification of outcomes at seven levels. The first level is the resources expended by the project (1). These resources produce activities (2), which involve people with certain characteristics (3). These people will have reactions to their experience (4) that can change their knowledge, attitudes, skills and/or aspirations (5). If these changes happen, then people may change on-farm practices (6), thereby achieving an end result of economic, social and/or environmental value (7). Level 7 represents the ideal ultimate outcome.

Bennett's Hierarchy provides a framework for collecting the evidence required to tell evidence-based performance stories that link activities to intended outcomes, similar to those advocated by Mayne (2001).

Two months after the Target 10 project as a whole had been criticized for not understanding its impact, staff produced the first evaluation report using Bennett's Hierarchy as a framework and drawing on existing data supplemented by a small survey. Target 10 was able to tell its performance story underpinned by a body of credible evidence.

A useful minimum requirement for evaluation had been established but, while an evidence-based performance story had provided a coherent story of impact, it was recognized that it didn't provide a full explanation of 'what worked for whom and why'. This became a key evaluation question. Working with a market research company to establish market segments, and collecting data internally, a number of patterns were identified and analysed (McDonald, 1997; McDonald and Rogers, 1999; Rogers and McDonald, 1999) according to the realist framework of context–mechanism–outcome configurations (Pawson and Tilley, 1997; Julnes et al., 1998).

While project staff used these additional insights about their clients to improve their project, this work also demonstrated how internal and external efforts could be combined to build a better understanding of the different causal paths. Lastly, it demonstrated the value of commissioning and undertaking research into and trials of different evaluation methods.

Reflection and Replanning

By the end of this phase, two developments were encouraging the expansion of evaluation to other projects. Both Target 10 evaluation reports (the initial Bennett's Hierarchy performance story, and the more detailed analysis of what worked for whom and why) had received external validation through independent blind critical review.

One of the project team who had promoted evaluation and experienced the benefits of it then moved to a senior position of responsibility as Executive Director, Agriculture. This gave significant momentum to the adoption of evaluation within the Division.

A set of principles and a framework for developing an evaluation plan had

been documented during this phase (McDonald et al., 1997) and were used to guide the expansion. The principles made explicit: the role of evaluation as a change agent; the need to plan evaluation during project development; the need to be utilization-focused (Patton, 1997*); and the need to use a variety of methods as appropriate. A modified version of the framework is summarized later in this article (see Box 1).

Phase Two: Experimenting with Volunteer Projects (18 Months)

The second phase, a piloting phase, involved developing evaluation strategies across a range of projects to test whether the approach taken in the first phase would hold in different fields of activities and to determine what modifications might be required. All were large projects with budgets in excess of AU$1m per year; most comprised a variety of sub-projects.

Typically, a member of the evaluation support team established for this purpose would work with project teams to develop a programme theory, key evaluation questions and a plan to answer those questions. Project teams would then manage the rest of the evaluation, contacting the support team when they needed assistance. During this phase and subsequent phases, how projects decided to conduct their evaluations varied. While some projects undertook all the evaluation in-house, others contracted out data collection and interpretation. Some projects called in external experts as mentors, or submitted their work for external expert review. Issues such as the size of the project, the maturity of the project and the requirements of stakeholders determined the approach chosen. If necessary the evaluation support team would facilitate specific training programmes or work with projects when they encountered difficulties.

Three important findings emerged during this phase. The first was the importance of encouraging methodological diversity. When working with these projects, the methodological challenges that emerged often had no clear answer. Project teams were encouraged to experiment with different approaches before making a final decision. For example, staff in the Farm$mart project, which trained farmers in business planning, had differing opinions about which data collection methods should be used to gather information about impact. It was therefore decided that each of the six regions would try a different method of gathering data about farmers' behaviour change. In each case, the evaluation support team provided additional internal or external support. At the end of the trial, the project team was able to make an informed decision on which combination of approaches to use.

Another finding was the benefit of encouraging short learning loops, where the results of short evaluations were used during implementation for formative purposes. For example, the GrapeCheque extension sub-project, which encourages grape growers to adopt improved viticultural practices, was using Goal Attainment Scales (Kiresuk et al., 1994) to track adoption of a particular technology. After 12 months, adoption rates had not met expected targets. Group dialogues were held to identify the reasons for non-adoption, using ORID

(Objective–Reflective–Interpretive–Decisional) procedures (Stanfield, 1997). Results revealed that the information being provided was too complex, so research scientists were called in to streamline the content of the extension package. After revision, GrapeCheque went on to exceed its adoption targets. These short learning loops provided early benefits for projects, and increased project commitment to evaluation generally.

The third finding was the need for a modified programme theory to underpin the evaluation of agricultural research projects. The impact of research is typically felt long after a project has finished and usually depends on the actions of one or more third parties. It is neither reasonable nor practical to hold research projects accountable for final impacts that are influenced by external factors and are unlikely to be fully achieved during the life of the project. The evaluation support team worked with project teams, with advice from Dr Claude Bennett, the originator of Bennett's Hierarchy, to revise the generic programme theory to make it more appropriate for research projects. Projects were encouraged to use this adaptation of Bennett's Hierarchy to articulate how the project would contribute to the greater goals of the agency and government policy, then identify how far up the causal chain they could provide useful information. In most instances research projects were expected to hypothsize an intended causal chain, but be accountable for delivering outputs (new knowledge or tools) and for how well these were placed in the market, including the response of the immediate recipients of the research findings.

Reflection and Replanning

Assessment undertaken at this stage found that the major benefit of evaluation to project teams was in assisting project design and developing a common understanding of the programme theory behind the project. It was too early for respondents to comment on the utility of the impact evaluation work in which they were engaged.

Around this time, the Division received funding for a major new initiative based on innovative science and technology. One of the reasons for the success of its funding application was that the Division could demonstrate its evaluation capability. Investment in evaluation was starting to deliver financial benefits. Part of the guarantee given by the Division in seeking this funding was that there would be credible evaluation of each project funded under the initiative as well as evaluation of the initiative as a whole.

Three options for ensuring adequate evaluation were considered. The first was to provide more funds to outsource evaluation efforts but this would have been expensive and unlikely to provide enduring benefits such as increased staff skills and organizational learning. The second was to develop an internal evaluation unit that would do the evaluation tasks, but this too was deemed unlikely to deliver a significant improvement in organizational learning. The option selected was to develop the evaluation capability of project teams to manage internal and external evaluations.

At this point, the explicit purpose of the capability building project was to enable projects to generate robust evidence of impact. This evaluative purpose

was to be complementary to any mandatory performance measurement requirement, and to ex ante economic assessments and investment decisions about the relevance of the project. In hindsight, having this aim helped address the concern, noted by Mayne et al. (1999), that 'a major challenge for institutionalized evaluation is to adequately address impact and continued relevance issues'.

Phase Three: Mandatory Evaluation for All New Projects (Two Years)

Evaluation became mandatory for *all* new projects, not just those involved in the new initiative. Approval for new project proposals was only given if a credible evaluation plan was provided. The message was clear: no evaluation, no money.

This decision may seem controversial, given previous discussions of the undesirability of making evaluation mandatory. The Division, however, was under a significant obligation to ensure that evaluation was planned and implemented for all new projects. From a pragmatic point of view, the Division believed that if it left evaluation to grow on a voluntary basis, it risked not fulfilling its contractual obligations to external funders. So it made evaluation mandatory but incorporated support and incentive strategies to ameliorate the possible negative effects of mandating evaluation. In particular, there was considerable scope within divisional requirements to make the evaluations useful for projects themselves.

By this time, the previous two stages had demonstrated to staff that there was little to fear from evaluation. Where projects had not delivered despite people's best endeavours, and the evaluation findings had been negative, staff had not lost their jobs or been penalized. Staff in these situations were given the opportunity to redesign their work or engage in different work. Although evaluation was mandatory, staff could move forward with confidence, using evaluation to research those components of their work they were fearful about. As a minimum, projects were required to provide an evidence-based performance story based on a programme theory with emphasis on impact assessment, customer satisfaction, and continuous improvement and science quality.

These requirements were part of a broader strategy to improve project results. Project teams were required – and given resources – to undertake a more thorough project development process, sometimes requiring up to a year and in partnership with stakeholders. During this time, communication and risk management strategies were to be developed as well as evaluation plans.

As the number of projects evaluation staff had to work with increased dramatically, a highly specific training programme was introduced. Consultants from universities and market research companies were initially engaged to assist in developing and delivering curricula, but the programme is now delivered by the evaluation support team in conjunction with a professional facilitator. The training is based on an evolving framework, summarized in Box 1, that focuses on five sequential tasks with a sixth task being overall management.

This framework was introduced through five days of residential training, delivered in two sessions separated by about a month. The first part focused on the first two issues (describing and framing) and the second part on the others. After

Box 1. A Framework for Developing Project Evaluation

- *Describing*: developing a shared understanding of the project, including a statement of what success looks like, and an agreed programme theory (usually Bennett's Hierarchy) of how the programme activities are intended to contribute to the desired outcomes.
- *Framing*: deciding the purposes of the evaluation, the audience, the specific evaluation questions to be answered, and the resources required.
- *Designing*: given the framing of the evaluation, deciding how the evaluation questions should be answered – taking into account sampling, data collection and data analysis.
- *Implementing*: gathering, analysing, interpreting and reporting data.
- *Using*: undertaking formal and informal processes to encourage use of the information from the evaluation and to encourage learning from the process of undertaking the evaluation.
- *Overall management*: applying sound project management principles to these activities as with any other activity, including milestone management, financial management, review, roles and responsibilities.

Note: Of these six tasks, only (3) Designing and (4) Implementing can realistically be contracted out. Project managers need to keep control of the rest.

several cycles of training, a third part was added – a six-month review based on learning contracts negotiated with participants at the end of the second part. This was designed to encourage implementation.

To further support projects, the evaluation support team also worked with project managers, developing and delivering a special two-day training programme for them. It was important that these managers were able to positively support their team members who were responsible for evaluation.

The emerging strategy for this phase drew on insights from agricultural extension that included the model for change behaviour that had been successful in Target 10 (Boomsma et al., 1996). This model hypothesized that achieving change required attention to four components: an educational opportunity, peer support, individual consultation to address difficult issues, and a community supportive of change.

Senior management demonstrated clear support for evaluation. The messages from the top emphasized the importance of being able to conceptualize a programme theory, collect relevant and credible evidence, and communicate succinctly. If you travelled in the elevator with the executive director, you were expected to tell a convincing evidence-based performance story before the elevator travelled from the 15th floor to the ground. If you couldn't do this elevator pitch, then you remained in the elevator until you could, despite people from other businesses coming and going. Word spread quickly amongst the staff (some of whom developed strategies for not going in the elevator with the executive director); the power of that action being seen as evidence of senior management's powerful and very visible commitment to evaluation.

Experts in particular areas of evaluation were invited to work with the evaluation support team on various activities including internal workshops and public

forums. Topics have included using Bennett's Hierarchy to integrate research and extension, the limitations and possibilities of performance measurement, realistic evaluation and demonstrating cause-and-effect relationships.

Reflection and Replanning

The review of this phase drew on data gathered from participants individually and as a group during training through a variety of methods: photo elicitation, sociometry, the daily 'dartboard' (a visual analogue of ratings of various elements), a daily debrief by the training team, and written questionnaires at the end of the training. Additional data were collected through an internal evaluation that involved:

- structured telephone interviews with participants from the first four training programmes;
- semi-structured interviews with 10 project managers involved in the early stages (who had also been interviewed at the end of Phase Two); and
- semi-structured interviews with five senior investors and three external experts who were familiar with the work of the evaluation support team (Dart, 2002).

This evaluation also drew on the records and observations of the evaluation support team.

A summary evidence-based performance story based on the above evaluation

With funds from the Victorian Government and knowledge based on piloting work, the evaluation support team has:

- developed and delivered a tailored change behaviour programme with formal training and after-training strategies;
- developed a framework for evaluating research and its dissemination;
- conducted research on impact evaluation and other issues;
- debated its lessons and experiences with international experts and invited other experts to work with the Division's staff; and
- identified a change model that describes the process of building an evaluation capability in projects.

Approximately 18 percent of Agriculture Division staff have attended training or participated in the early trials. Other divisions, and, occasionally, external agencies, are now participating. Almost three-quarters of key project managers have undertaken formal training, as have all, but the most recently appointed, internal industry investors.

However, while the evaluation support team had focused on changes at an individual level, the review showed gaps in supporting these people when they returned to their teams and were required to implement evaluation within the whole-of-project context.

In general, the reaction from staff, both investors and providers, has been positive despite the fact that evaluation is mandatory.

> When you start evaluation first there is fear, then there is discovery. This leads to confidence and empowerment.
>
> Research scientist

Approximately three-quarters of programme participants have used their evaluation findings to demonstrate impact, alter project strategies or develop/redevelop their project plans. Eighty percent have used project logic to plan new projects or report results, or to enable them to think more broadly about their project.

In two and a half years, the evaluation support team has influenced the culture of the Division, facilitating a pathway to organizational learning. Many people are thinking more critically about how their projects achieve results, questioning the logic and rationale of their actions, and, in the process, improving what they do.

> It effected cultural change. Only now has it gelled with me that I operate differently. Now we can question what we do as a matter of course. We have an atmosphere of critique. We are actively seeking how we can do better. We didn't have that before.
>
> Key project manager

> I do a lot of work with scientists, who have got a great science project and you know [their attitude is] 'Just trust me I am a scientist. I know what needs to be done.' The discipline of them having to do a project plan that incorporates an evaluation plan, I think is one that helps them actually sit back and critically analyse where their project fits . . . into a bigger world.
>
> External observer

Internal investors and key project leaders are beginning to report increased success at attracting external funds and are attributing this to the improved planning, performing and reporting that has resulted from implementing the evaluation framework. The work of projects, including their own evaluations, is often subjected to external review, which provides confidence about the credibility of the work or provides insight for learning.

But it is not all good news. In some project areas there is confusion between investors and project teams regarding the demand for evaluation. When demand signals are weak, projects tend to emphasize formative evaluation. When demand signals are strong, projects focus on impact evaluation. Without strong demand signals, the effect noted by Mayne et al. (1999) is evident: impact evaluation becomes a challenge and project evaluation becomes more focused on in-house issues.

> And I have had no feedback whatsoever. I got up to a stage where I got sick of them [investors] saying, "Have you done it? Where is it?" This was six months after I had submitted it.
>
> Key project manager

> . . . [we have] a better acceptance [of evaluation] at the key project manager level. I am not sure how far that goes down in the organisation. I am not sure how many subprojects are on board with evaluation.
>
> Internal investor

In response to this finding, the evaluation support team has developed a strategic plan to position itself at this crucial interface between investors and providers in order to work on balancing supply and demand.

Also, some people are yet to be convinced and are slow to adopt evaluation. Reasons for non-adoption are usually complex, and the evaluation support team plans to seek a better understanding of why this is occurring.

Another issue to emerge is that when projects reach a certain level of capability they become independent of the evaluation support team and move outside the framework. While this is not necessarily a problem, on some occasions a bigger and better evaluation has become a goal in its own right, outstripping demand. In these cases questions of utilization are left behind. Like a cat dropping a 'dead mouse at the doorstep', these efforts produce reports that no one wants to pick up.

Phase Four: Expansion and Consolidation (Current Phase)

Our vision for the future is that a small team of people will maintain the Division's evaluation capability and incorporate new information about evaluation into the system. Meanwhile two or three PhD students will research evaluation issues that add to the body of knowledge.

The Division is receiving an increasing number of unsolicited requests for advice and information about what is happening, both from other divisions and external agencies. There is potential to work more broadly outside the Division, and to include external agencies. In some senses this could be seen as an extension of the 'pincer movement' – of working simultaneously with individual projects and senior decision makers. It may not be possible for an agency to build its evaluation capability without congruent frameworks and incentives being used by the larger organization and funders. In order to address this, the evaluation support team has recently been asked to develop an options paper for developing an evaluation capability across all ten departmental divisions.

Our journey is not complete. There are still challenges to be resolved, but there is sufficient evidence to suggest we have made progress and it is worth continuing. As one project manager said, 'I think it is extremely hard to bring about cultural change. Saying that, if the aim is to move the ship by 90 degrees, then I think they have moved it 45 degrees!'

Emerging Themes

Seven major practical themes, drawn from this case study, are offered as overarching recommendations for building an evaluation capability and are discussed below.

Think Big, but Start Small – Stage, Trial and Grow Evaluation

The gradual growth of evaluation through a series of implementation, review and redevelopment stages, as in our case, produces benefits in several areas. This is a practice that Preskill and Torres (1999b) have recommended. Risks are reduced

by trialling and testing before going to full implementation across all projects. This makes methodological experimentation possible. It also generates real, local examples within the context of our agency for use in training, demonstrating the feasibility and utility of methods while highlighting practical issues and strategies and providing contact details of people to help with data collection or analysis. It allows the agency time to develop its external networks and resources of consultants and advisers, and to develop its internal networks of expertise, trusting relationships and knowledge of internal resources. It also allows time to gain acceptance that there is no 'one suit fits all' evaluation strategy. In terms of the model of behaviour change, it allows time to develop a community supportive of change. It takes the fear out of evaluation. Other agencies may move forward more quickly if they decide to standardize early on, but may not have the perceived legitimacy for such a decision without the work of trialling and reviewing alternatives.

But there is a perceived downside to this staged approach. Several agencies who have approached us for advice have claimed that they cannot afford to take this long to build their evaluation capability. It may be possible to achieve similar results in less time than the five years it has taken us, but it may also prove impossible to proceed much quicker and still be effective. McAllister (1998: 107) has argued that we need a long-term perspective for building evaluation capacity. She states that '. . . however we do it, we will have to be in for the long-run – at least eight to ten years'.

This should not, however, be read as an argument for a luxurious timeline without visible results. It is very important to quickly and repeatedly get 'points on the board', to be seen as responsive to short-term needs, and of some use immediately.

Address Both Supply and Demand

When building evaluation capability, there is often a temptation to focus exclusively or excessively on the supply side. Developing and running training programmes, for example, is a highly visible activity, and can provide positive reinforcement for evaluation staff. However, without adequate attention to demand – understanding actual demand (as opposed to espoused demand) and the demands of different stakeholders – these efforts may ultimately be futile and lose credibility. As Dr Claude Bennett advised the Department of Natural Resources and Environment (Bennett, 1977, 1999):

> I see a major challenge to the continuation of the momentum of your evaluation program. . . . As overall evaluation efforts widen, there are associated risks. If project staff begin to believe their completed evaluations are ignored/not acted upon/misused, morale associated with conducting evaluations will plummet.

Building an effective evaluation capability requires balancing supply and demand and iteratively raising the level of each. For example, a demand to account for expenditure can be simply met, but to achieve goals such as creating a learning culture and robust impact evaluation methods requires a capacity to produce more comprehensive information and to use this information.

Possibly because much of the work in this case study has focused on supply, problems with utilization have become apparent. Where projects have been successful in matching supply and demand, they typically instigate formal or informal partnership arrangements with key stakeholders. It becomes possible to manage changes in the demand for evaluation. One consequence is that any summative evaluation is more readily understood and acted upon.

At times the demand for impact evaluation can be driven by communication goals and there is a danger that this will bias the evaluation towards seeking out evidence of positive outcomes. Occasionally this has proven difficult to manage.

Use a Pincer Movement, Working Top-down and Bottom-up

The conscious strategy has been to take top-down *and* bottom-up approaches. A top-down approach involves working with senior managers to develop their support for evaluation and to understand and meet their evaluation needs, mandating evaluation across projects and providing central resources. A bottom-up approach involves working with project staff and managers to understand and meet their evaluation needs, and developing policies and frameworks from specific cases. To take a top-down approach only might result in a token attention to evaluation; a solely bottom-up approach would take a longer time, and might produce many evaluations that do not focus on the agency's core business.

Discussions about whether to make evaluation mandatory sometimes proceed as if top-down and bottom-up approaches are mutually exclusive – as if making evaluation mandatory would inevitably lead to token compliance. In our case, external pressure required prompt and comprehensive action in making evaluation mandatory, and possible ill effects had to be ameliorated through simultaneously attending to bottom-up processes.

Address Different Aspects of Behaviour Change

Milstein et al. (2002) have commented that one of the greatest barriers to evaluation is fear. They state that if:

> experience doing evaluation inherently builds capacity, then fear and misconception, which prevent projects from getting under way in the first place, stand as the greatest obstacles to progress. Common concerns about evaluation must be addressed and predictable lines of defence countered to overcome the inertia associated with beginning new evaluation projects. . . . the perception of a punitive environment is antithetical to sound evaluation practice. To build capacity, agency leaders must convince staff and stakeholders that answering evaluation questions thoroughly and honestly is more important than getting results that look and feel good. (Milstein et al., 2002: 42)

From the beginning of this work, the executive director and the evaluation support team were keen to remove fear. Some of the fear was removed by using the model of change previously mentioned (Boomsma et al., 1996) and by having the support team there to address difficulties.

One important component in the model of change addressed creating a community supportive of change through organizational behaviour. For example, to date, no one has been penalized if evaluations showed that their projects failed

despite their best efforts at delivery. Also, the executive director always gives feedback to project teams when he receives reports, an award for learning in evaluation has been instigated, and staff are supported to attend evaluation conferences and courses. These actions start to address Davidson's (2001) point about the important role of symbolic actions in developing an evaluation capability.

During the evaluation of Phase Three it became clear that a second model of change was operating, this time focused on how evaluation capability emerges in projects. This new model described a journey. Initially, the project manager and one or two team members learn about evaluation, then trial it in a small way. If this is a positive experience, they endeavour to engage the whole project team to undertake whole-of-project evaluation. Once the team has mastered this and has received the benefits of doing evaluation, it may look to improve and strengthen its evaluation capability (Dart, 2002). By not having articulated and used this second theory of change at the project level, the evaluation support team missed some valuable opportunities to strengthen its impact.

Develop a Common Evaluation Framework

While many organizations around the world have found programme theory useful in evaluation (Bickman, 1990; Rogers et al., 2000), this case study found it particularly helpful to use a generic programme theory, in our case Bennett's Hierarchy, as the basis for project evaluations. Initially we allowed projects to choose from several forms of programme theory but most opted to use a free-form approach using box and arrow diagrams. However, they tended not to move on but instead kept refining and complicating the picture. In these cases, programme theory had lost its value as a learning tool.

However, one form was delivering benefits. Those projects that had used Bennett's Hierarchy were making significant gains. It was easy to use and quickly gave projects a causal model and insight into their work. Even using existing data, most projects were able to write a performance story and quickly identify any gaps in their knowledge. The hierarchy also provided a common language for projects to discuss their work, supporting planning and communication efforts as well as helping to frame evaluations. It also provided a means for aggregating data across projects.

Given this evidence, Bennett's Hierarchy was strongly recommended as the preferred theory-of-action approach, but needed to be adapted for research projects (as mentioned previously).

However, project logic on its own is insufficient to complete the evaluation task. There is a risk that projects will excessively focus on intended outcomes and average outcomes and develop only a weak case for causal attribution (Rogers and McDonald, 1999). We therefore encourage projects to look for unintended impacts by considering economic, social and environmental accountabilities and using negative programme theory (Weiss, 1998). We advocate going beyond average outcomes to identify sub-groups and multiple pathways through projects to get some insight into the crucial matter of what worked for which groups and why.

Finally, where appropriate, we encourage projects to increase their evidence for causal attribution beyond simply measuring performance against a programme theory. This can be achieved by considering issues such as:

- counter-factuals (which look at how a specific outcome changes with the presence or absence of a cause such as a particular programme);
- underlying generative mechanisms; and
- the combination of ingredients that must be present for a particular cause to have an effect (Mark et al., 2000; Mark, 2002).

From a pragmatic point of view, it is often sufficient to provide evidence that the hypothesized programme theory is working, particularly if causality has been established from some prior evaluative exercise. It is not feasible, efficient or sensible to undertake a full-blown evaluation every time (Mark et al., 2000). However, it is important to have advanced evaluation tools available when required.

Build Knowledge about What Works in Evaluation in your Context

From the beginning of this project, the Division has had an ongoing commitment to build knowledge about evaluation. Some of these are the expected activities of professional staff development: completion of graduate courses in evaluation; attendance at evaluation conferences and workshops; and engagement of consultants on various projects as part of knowledge and skills development.

In addition, this agency has taken the unusual step of framing evaluation as a research and development activity. In the same way that world-class scientists are often invited to spend time with the Division to build up the Division's science capability, experts in various aspects of evaluation have been asked to advise us on evaluation projects. The Division also commissions and conducts research and trials into evaluation techniques.

The evaluation support team has also developed ongoing linkages with other internal and external groups to work collaboratively on research into impact-evaluation issues and techniques, including fieldwork and trials.

Why is this important? We believe that knowledge about evaluation – how to do it better, and how to develop evaluation capability – can best be built through partnerships between government and research institutions. This provides a culture that encourages exploration of difficult issues and leads to a symbiotic relationship where theory informs practice and practice informs theory. It provides a capacity for critical inquiry. Given that there is a research culture within the Division, introducing evaluation research was not a difficult step.

We also believe that good evaluation practice depends on context. What works well for one organization in doing evaluation and building evaluation capability may not necessarily be appropriate in another organization, and needs to be carefully assessed and trialled.

Walk the Talk: Systematically and Visibly Evaluate Each Stage

If the expectation is for people to change the way they go about their business then those who are promoting that change must themselves demonstrate that

behaviour and be subject to the same requirements. The evaluation support team complies with its own evaluation plans and results and makes these openly available. We use this information to develop our communication performance stories for funding bodies and to redesign our capability-building programmes as necessary.

Conclusions

This article has outlined seven practical recommendations for building evaluation capabilities and provided information about the context for developing these recommendations. We look forward to learning from others about the extent to which these recommendations might be applied to other situations, and about other issues and strategies in building evaluation capability.

References

Bennett, C. F. (1977) *Analysing Impacts of Extension Programs*. Washington, DC: USDA Extension Service.

Bennett, C. F. (1999) Personal Communication. [author: sage OK?]

Bennett, C. F. and C. Rockwell (2000) *Targeting Outcomes of Program (TOP)*. Available from: http://deal.unl.edu/TOP/synopsis.html.

Bickman, L., ed. (1990) 'Advances in Program Theory', *New Directions for Program Evaluation* 47. San Francisco: Jossey-Bass.

Blalock, A. (1999) 'Evaluation Research and the Performance Management Movement: From Estrangement to Useful Integration?', *Evaluation* 5(2): 117–49.

Boomsma, J., B. McDonald and M. Walton (1996) *Target 10: The First Three Years*. Melbourne: Department of Natural Resources and Environment, Government of Victoria.

Dart, J. (2000) *Stories for Change: A New Model for Agricultural Extension Projects in Australia.* Melbourne: PhD thesis, Institute of Land and Food Resources, Melbourne University.

Dart, J. (2002) *Evaluation of the Evaluation Support Team*. Unpublished report.

Davidson, E. J. (2001) 'Mainstreaming Evaluation into an Organization's "Learning Culture"', Paper presented at the Meeting of the American Evaluation Association *Mainstreaming Evaluation*. St. Louis, 8 November. Available from: http://unix.cc.wmich.edu/~jdavidso/olcult.ppt.

Davies, I. (1999) 'Evaluation and Performance Management in Government', *Evaluation* 5(2) 150–9.

Department of Natural Resources and Environment (2001) *Innovation: Sustaining Export Growth in the Food and Agriculture Sector.* Melbourne: Government of Victoria.

Fetterman, D. M. (1996) 'Empowerment Evaluation: An Introduction to Theory and Practice', in D. M. Fetterman, S. J. Kaftarian and A. Wandersman (eds) *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*, pp. 3–46.Thousand Oaks, CA: Sage.

Julnes, G., M. Mark and G. Henry (1998) 'Promoting Realism in Evaluation: Realistic Evaluation and the Broader Context', *Evaluation* 4(4): 483–504.

Kiresuk, T., A. Smith and J. Cardillo, eds (1994) *Goal Attainment Scaling: Applications, Theory and Measurement.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Leeuw, F. L., J. Toulemonde and A. Brouwers (1999) 'Evaluation Activities in Europe: A Quick Scan of the Market in 1998', *Evaluation* 5(4): 487–96.

Mark, M. (2002) *What Works and How Can We Tell?* Melbourne: Department of Natural Resources and Environment, Government of Victoria.

Mark, M., G. Henry and G. Julnes (2000) *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs*. San Francisco: Jossey-Bass.

Mayne, J. (2001) *Performance Measurement.* Melbourne: Department of Natural Resources and Environment, Government of Victoria.

Mayne, J., S. Divorski and D. Lemaire (1999) 'Locating Evaluation: Anchoring evaluation in the executive or the legislature or both or elsewhere?', in R. Boyle and D. Lemaire (eds) *Building Effective Evaluation Capacity*, pp. 23–52. New Brunswick, NJ: Transaction Publishers.

McAllister, E. (1998) 'Overview and Conclusions', in K. Mackay (ed.) *Selected Proceedings of the World Bank Seminar: Public Sector Performance – The Critical Role of Evaluation*, pp. 121–5. Washington, DC: World Bank.

McDonald, B. (1997) *Improving Dairy Cow Nutrition Using a Marketing Approach.* Melbourne: Department of Natural Resources and Environment, Government of Victoria.

McDonald, B. and P. Rogers (1999) 'Differentiated Program Theory Evaluation as an Analogy to Market Segmentation', Paper presented at the Annual Meeting of the American Evaluation Association, Orlando, FL.

McDonald, B., J. Winston and J. Boomsma (1997) 'Getting Started in Evaluation Extension: A Case Study of Practices and Principles Used in Target 10', in *Proceedings of the 2nd Australasian Pacific Extension Conference 'Managing Change – Building Knowledge and Skills'*. Wodonga, Australia: APEN.

Milstein, B. and D. Cotton (2000) 'Defining Concepts for the Presidential Strand on Building Evaluation Capacity', Evaluation 2000 Conference Information. Available from: http://www.eval.org/eval2000/public/presstrand.pdf

Milstein, B., T. Chapel, S. Wetterhall and D. Cotton (2002) 'Building Capacity for Program Evaluation at the Centers for Disease Control and Prevention', in D. Compton, M. Baizerman and S. Stockdill (eds) *The Art, Craft and Science of Evaluation Capacity Building, New Directions for Evaluation* 93, pp. 27–46. San Francisco: Jossey-Bass.

O'Brien, G. and G. Hepworth (1994) *Evaluation of a Pasture Management Program for Dairy Farmers in West and South Gippsland.* Melbourne: Department of Agriculture, Government of Victoria.

Patton, M. Q. (1997) *Utilization Focused Evaluation,* 3rd edn. Thousand Oaks, CA: Sage.

Pawson, R. and N. Tilley (1997) *Realistic Evaluation.* London: Sage.

Preskill, H. and R. T. Torres (1999a) *Evaluative Inquiry for Learning in Organizations*. Thousand Oaks, CA: Sage.

Preskill, H. and R. T. Torres (1999b) 'Building Capacity for Organizational Learning Through Evaluative Inquiry', *Evaluation* 5(1): 42–60.

Rogers, P. and B. McDonald (1999) 'The Three Achilles Heels of Program Theory' Paper presented at the Australasian Evaluation Association Conference, *Evaluation: Challenging Boundaries*. Perth, WA; 6–8 October.

Rogers, P. J., A. J. Petrosino, T. Hacsi and T.A. Huebner, eds (2000) *Program Theory Evaluation: Challenges and Opportunities, New Directions for Evaluation* 87. San Francisco: Jossey-Bass.

Russon, C. and M. Patel (1999) 'Evaluation Capacity in Eastern and Southern Africa', *Evaluation News and Comment* 8(1): 14–5.

Scriven, M. (1997) 'Book Review: Empowerment. Evaluation: Knowledge and Tools for Self-Assessment and Accountability', *Evaluation Practice* 18(2): 165–75.
Stanfield, B., ed. (1997) *The Art of Focused Conversation*. Toronto: The Canadian Institute of Cultural Affairs.
Weiss, C. H. (1998) *Evaluation: Methods for Studying Programs and Policies*. Upper Saddle River, NJ: Prentice-Hall.
Williams, B. (2001) *Building Evaluation Capability*. Unpublished paper. Available from: http://users.actrix.co.nz/bobwill/.
Wissemann, A. (1996) Personal communication. [author: Sage, OK?]
World Bank (1994) 'Building Evaluation Capacity', *Lessons and Practices Paper No. 4*. Washington, DC: World Bank.

BRON McDONALD is the project manager within the Agriculture Division of the Victorian Department of Natural Resources and Environment, Australia, responsible for the development of an evaluation capability. Please address correspondence to: [author: please supply].
[email: bron.mcdonald@nre.vic.gov.au]

PATRICIA ROGERS is an evaluator, researcher and educator based at CIRCLE (Collaborative Institute for Research, Consulting and Learning in Evaluation) at the Royal Melbourne Institute of Technology. Please address correspondence to: CIRCLE, RMIT University, Level C, Building 1, 124 Latrobe St., Melbourne 3000, Australia. [email:patricia.rogers@rmit.edu.au]

BRUCE KEFFORD is the Executive Director of the Agriculture Division of the Victorian Department of Natural Resources and Environment, Australia, and has overall responsibility for the delivery of services to agricultural industries. Please address correspondence to: [author: please supply].
[email: bruce.kefford@nre.vic.gov.au]