Promoting Organisational Learning via Evaluation: the New Accountability?

John M Owen
Centre for Program Evaluation
The University of Melbourne
j.owen@unimelb.edu.au

Paper presented at the bi-annual conference of the European Evaluation Society, Seville, Spain, October 2002
Introduction

This paper focuses on meanings of accountability and the use of evaluation to meet accountability needs in social and educational systems. I am concerned here with multi-level and multi-site interventions designed to have a widespread impact on a specified community. Typically such a system involves the development of policy at a central level, and the implementation of this policy at a range of sites.

While this is a common pattern of service provision, the integration of evaluation into these arrangements has received scant attention by theorists and leaders of evaluation practice (Herrel and Straw 2002). Recently there have been several ineffective large-scale evaluations in Australia, which would have benefited by a greater conceptual understanding of the complexities associated with multi-level program provision.

There is no doubt that accountability is currently a ‘buzz’ word in the public domain. The recent collapse of large corporations in the United States, Australia and elsewhere has led to an erosion of confidence in the quality of financial accounting standards and the behaviour of senior executives. The very fabric of economic life in western societies has been called into question. Governments are now introducing measures designed to restore confidence in the stock market, and to punish company directors and others who have attempted to defraud their shareholders.

Concern for accountability is also current in the public sector. There is evidence that the public needs to be satisfied that spending on public sector programs and institutions is both effective and efficient (Gilbert 1998). Accountability is a catch cry in today’s public sector politics. It is one of the slogans used to lobby for more and better evaluation.

However, there is a danger that strident calls for accountability can lead to systemic arrangements which can have deleterious effects on the quality of programs and the clients of these programs, resulting in outcomes which are the opposite to those they are designed to achieve (Earl and Lafleur 1997).

Here, I am concerned with the use of different arrangements by which evaluation can generate knowledge to fulfil accountability requirements and ensure that programs are effective. Effectiveness questions include the following:

- how can we assure that policy/program interventions meet the needs of the intended recipients, and
- how can evaluation be structured to ensure that this outcome is achieved?

For this discussion to be meaningful, we need to consider these arrangements within the ‘realpolitik’ of the current situation regarding accountability, which
has just been discussed. It is a truism that government and other funding bodies provide the financial support for social and educational programs. Spending on programs often involves the allocation of considerable resources drawn from the public purse, and there is a need for those involved to provide evidence that the funds have been spent wisely.

**Stakeholders and Knowledge Needs**

Before we can discuss accountability, we need to identify key players by which I mean those program stakeholders with a direct interest in the intervention being examined. For most large scale interventions key players are the:

- funders: the source of resources, for example a government treasury or an organisation independent of government which has its own source of funding, for example a philanthropic trust
- auspicers (or sponsors): the organisation which develops or is responsible for the policy under which funds are disbursed, for example a government department
- providers: those who deliver interventions under the auspicing or sponsoring organisation policy, for example a school or school district, a town council
- clients: those who receive the program and are expected to benefit from it, for example school students or town residents.

There may be additions to this list, for example, when there are multiple funding sources. It is also possible that the funding source and the auspicing agent are part of the same organisation, as is the case for the World Bank. In some international development scenarios, the government of countries, where providers are located can be involved at all stages of policy development and delivery, and are legitimate stakeholders.

My view is that all key players need to have relevant knowledge or information about a given intervention as a basis for ensuring that the intervention proceeds smoothly and meets the needs of the program recipients. In fact the definition of evaluation I have developed is couched in these terms:

*Evaluation is the process of negotiating an evaluation plan, collecting and analysing evidence to produce findings, and disseminating the findings to identified audiences for use in describing or making decisions related to the intervention or evaluand (Owen, 1999: 4).*

We know that there was a time when policy and program decisions were based on the best judgements of people in authority. It was assumed these people used a combination of privileged knowledge of the context, political wisdom,
professional training and logical analysis. Systematically assembled information played almost no part in their decision making.

During the past decade, this situation has changed and the use of evaluation findings is becoming the watchword in the current wave of public sector reform. An issue we need to discuss here, and which should continue to be a point of debate concerns the use of findings which result from systematic data collection and analysis.

In the context of multi-level evaluations, the following points need to be made explicit:

- the primary purpose of policy development is effective program provision
- there are at least three levels of stakeholders who require knowledge about the program (funders, policy makers, program deliverers)
- the knowledge needs of these stakeholder groups are likely to be different
- evaluation resources are finite (in my view about 10-15% of the budget at most)
- we need to make decisions about how these evaluation resources can be divided between key stakeholder groups
- the knowledge needs of all stakeholder groups will vary with time, that is they will vary as the program proceeds.

Accountability

Accountability has long been thought of as the obligation to report on the exercise of responsibilities conferred on managers for the expenditure of funds and implementation of mandated activities.

For this discussion, let me assume that one of the primary knowledge needs of the funding source stakeholder can be satisfied by financial information. This will, of course, vary from program to program but on the whole financial information would be in the form of budgets to accompany program plans, interim reports on program expenditure, and financial audits when the program has been completed.

It is clear that these kinds of information can be supplied without the need to undertake evaluations which are consistent with the definition I have supplied above. What is of interest in the rest of this paper are ways in which evaluation arrangements can supply information for accountability needs. To begin this discussion I would like to review some typical evaluation practices in multi-level systems, and begin with what I have labelled the Traditional Perspective.
Traditional Perspective: Emphasis on Accountability UP

Given that accountability is associated with reporting, when associated with evaluation it represents one kind of utilisation of evaluation findings (Owen, 1999). Traditionally, evaluation for accountability has been associated with reporting results such as:

- were targets achieved?
- were the goals met?
- were the objectives achieved?

In this perspective there is an emphasis on what I will call Accountability UP. Auspicers set the agenda, the methodological approaches to be used, and are the primary audience for the findings. Providers are required to collect and assemble evaluation data and report it to the auspicer.

Typically auspicers use the information for two purposes:

- aggregating the information from all providers to report to the funding agency, and
- for making decisions about the future fate of individual providers and key individuals responsible for the performance of an individual provider.

Little or no use of the evaluation information is made at the provider level, because it is not fed back or because aggregated information is irrelevant. In this perspective, providers are not evaluation users.

Traditional Approach: Use of Performance Measurement. One of the most pervasive trends in providing evaluative information for accountability purposes is the use of performance measurement (PM). This involves the measurement and reporting of processes and outcomes as part of the management of a given program. Usually, these measures are in the form of discrete quantitative indicators.

The use of PM has been championed by some well-known evaluators as a way of collecting and disseminating simple but understandable information, and as the basis for reporting up to the next level of management (Wholey, Hatry et al. 1994). The use of PM is consistent with the ‘payment by results mentality’ in private industry, where senior company executives are rewarded in accordance with the movement of the company share price.

On the other hand, the use of PM has been roundly criticised by some evaluation theorists (Perrin 1998). However, Perrin believes that PM can be used for monitoring a program, provided certain conditions are met.

These include that the indicators:
- measure important program aspects of program provision
• are used for monitoring purposes, not final assessments
• are tested and agreed to by program staff, and revised regularly
• are used as a set in order to examine a variety of program aspects
• need to be interpreted to make information meaningful.

In my experience, there are several concerns about the use of performance indicators in evaluation. One is that PM does not provide an interpretation of meaning; this is left to those who receive the performance indicator data set.

Another concern is the lack of a framework to enable different performance indicators to be linked. Such a framework would enable us to look for links between program components, for example, between process and outputs, outputs and outcomes. The use of program logic is one means by which these linkages could be made and I have used this technique in some recent studies we have undertaken (Owen and Lambert 1995). However, my experience is that the use of performance measures does not usually have this level of inclusion.

In most cases PM divorces cause from effect. PM provides information about outputs and outcomes and it is up to those who need this information to then make sense of what causes this pattern of evidence. This is usually information about processes or implementation. Implementation factors could be related to what has happened in the past, or what should happen in the future. While divorcing cause from effect may cause concern among social science theorists, it does provide those who receive the information with the flexibility to make decisions about elements of the program that fall under their control, and this may explain the popularity of the PM movement among program managers.

Other criticisms of PM as an approach to evaluation are that the information is sometimes too late to influence decision-making about the program. In addition, the use of pre-ordained outcome measures encourages a compliance mentality. This tends to focus staff on those outcomes which have been specified in advance, with the result that risk-taking is avoided, and adaptation of the program to meet client needs is discouraged. More importantly, there is the potential problem of goal displacement, measuring only what can be easily measured by quantitative measures. Finally, and in extreme circumstances, an emphasis on the collection and assembly of indicator information can divert staff from program provision issues to measurement issues (this is similar to the student testing syndrome found in some educational systems).
Alternative Perspective: Accountability UP and DOWN

I believe that there is an alternative ways for a system to implement accountability procedures. This relies on the adoption of a policy and evaluation schema which is more even handed in regard to accountability. This schema relies on the need for the auspicer to be accountable DOWN to providers in addition for providers to be accountable UP.

How would this more symmetrical approach to evaluation and program provision operate across a system?

For accountability DOWN the following elements need to be in place. The auspicing organisation would

- provide clear policy regarding the ways each provider in the system undertakes internal program development and evaluation.
- negotiate with all providers about what findings need to be reported UP to the auspicer
- provide the resources for providers to undertake meaningful and valid evaluations, and assistance to use the findings internally
- disseminate ‘usable evidence’ to providers. This could be in the form of indicators about the core business of the system which are more efficiently collected by the auspicer than by individual providers, yet can be useful to individual providers (Earl and Katz, 2002).

For accountability UP, the following elements need to be in place. All providers in the system would

- report on key outcome measures which are of use to the auspicer for their accountability to funder purposes
- show that they are using the findings from their individual evaluations for internal decision making and program improvement.

This approach to evaluation relies heavily on the ability of providers to undertake meaningful internal evaluations which lead to organisational learning. I think of such a scheme as being consistent with a ‘quality assurance’ approach. A quality assurance approach is based on auspicing organisations making it mandatory for the providers to incorporate evaluation into all stages of program provision.

A quality assurance framework is designed to serve the information needs, first and foremost, of the agencies delivering the program, and thus its major audience is internal to the program, managers and program deliverers. Accountability to the auspicing organisation comes by reporting that internal evaluation has taken place and that it is contributing to the effectiveness of program delivery on an ongoing basis.
A quality assurance scheme encourages learning throughout program delivery. One can think of evaluation being undertaken in one or more of the three zones suggested by Robert Brinkerhoff:

- **zone 1:** before the program commences.
  - what are the needs of the clients? what kind of program is needed? what is known about the program from other sources (such as research)?

- **zone 2:** during the delivery of the program.
  - has the program a defensible program logic which takes into account the findings of the before program analysis? (This is called an ex-ante evaluation by the World Bank)
  - is the program on-track? how do we need to refine the program to make it more responsive to the needs of the clients?

- **zone 3:** outcomes.
  - what objectives have been achieved? what have we learnt from this program?

In summary, a quality assurance scheme requires program providers to use evaluation information for all stages of program provision. For many providers this represents a sea-change in their approach to what they do and can be summarised by the following:

The use of a quality assurance scheme would mean that we would no longer think of providers only as knowledge distributors. Providers would also be centres of enquiry which is consistent with organisational learning. They would be creators as well as a distributors of knowledge. It has been said that a provider which is engaged in organisational learning is an agency characterised by a pervasive search for meaning in its work.'

**Implications for Auspicers and Providers**

Such a scheme cannot become routine without the support of auspicing organisations. They need to be committed to an expanded and somewhat different notion of evaluation compared to one based on performance measurement.

There are two key ideas here.

The first is that reporting for accountability purposes involves indications that a provider agency complies with the notion of evaluation to assist internal decision making, along the lines of what has just been described. That is, evidence of the use of evaluation in program decision-making must be supplied. This does not mean that the findings of a given evaluation have to be
supplied, but findings might be of interest to the auspicing organisation in a given case.

The second is that resources must be provided in program budgeting for internal evaluation. This also includes resources for training managers and other key staff about an approach to evaluation which will be new to many of them. There will also be a need for support for professional evaluators to work with program staff. I envisage that staff who work in these ways would have some knowledge about organisational learning and participatory principles which have been encouraged by Cousins, Patton, Torres and others (Cousins and Earl 1992), (Patton 1997), (Torres, Preskill et al. 1996).

My experience with procedures such as these is that, over time, an agency which sees the benefits of undertaking internally useful evaluations builds up an evaluation culture (Owen and McDonald 1999). This can lead to the desirable situation of creating a learning orientation among staff in the agency, an increase in professionalism and increased effectiveness and efficiency.

**Implementing the Alternative Accountability Perspective**

Innovations have been dogged by the adage that ‘it is fine in theory but it will never work in practice’. In the remainder of this paper I will introduce an example of this perspective which provides some indications of what is needed to implement an Accountability UP and DOWN evaluation framework.

School Accountability Framework in the State of Victoria, Australia. About a decade ago, a newly elected government set up a major program of structural, and curriculum reform which was designed to change the ways schools operated in the State of Victoria, Australia. While there were many reasons for this, the major one was that there was virtually no systematic data on the performance of the government school system in the State. The government directed that the state education department should arrange for periodic, independent reviews of school performance.

An Office of School Review (OSR) was set up to manage the implementation of what was titled an Accountability Framework.

All government schools were required to be involved in the development of a planning document, known as the school charter. This was a plan for the next three years of the work of the school. The charter was designed to enable schools to identify long term goals and implementation strategies. The charter involved significant work by staff, the Principal and the school council, which consisted of parents and teachers at the school.

In addition to the planning side of the charter, the school was also asked to develop methods for evaluating the impact of the objectives of their charter.
Each school was required to use the evaluation findings to continuously review the progress of their charter. Schools were also required to provide findings to the OSR. Examples included: the attendance record of students, and the results of surveys of parents, which tapped their opinions of the school.

A third initiative was that the OSR undertook state-wide tests of literacy and numeracy. In addition to this information being used by the OSR, the performance of each school was reported to each school.

The school thus had access to information which was used for modifying its school charter during the three year period of the charter.

This information was also used at the end of a three-year period to review the charter. At this time verification processes were carried out which involved external moderators, including university education faculty staff.

The original aim of the review was to verify that the school’s self assessment represented a fair statement of the school’s achievements over the three year period. In this sense the moderator played the role of an external inspector on behalf of the central authority. However, as the framework has matured, there has been an increased emphasis in the review process on support for the next round of charter development.

Thus, the triennial school review has taken on of an evaluation which encourages use for improvement of the charter and the school. This has been brought about by a combination of factors, including; the sympathy of most external moderators towards the information needs of the schools, and the provision of indicator information that was meaningful to principals and school review panel members. This included information which allowed them to compare the performance of the school with ‘like schools’, those with similar student populations on socio-economic grounds.

In summary, the School Accountability Framework now meets the information needs of both the auspicing organisation (the OSR) and individual schools, as follows.

The OSR had sufficient information to aggregate the findings to produce a system-wide (State level) picture on key indicators; achievement in literacy and numeracy, parental opinions of schooling etc, which could be examined both as a function of time and by school.

Each school had developed and evaluated its individual school charter plan and had used the information to develop a new three year charter. While there is no independent evaluation of this process, there is anecdotal evidence that
school principals have found these processes to be of high value in managing their schools.

Discussion

This paper has focussed on the need for mutual accountability involving auspicers and providers. It provides a scheme by which the information needs of all parties in a multi level system of program provision can be met.

In the case I have examined it was for schools to be accountable to an authority in the education department. In addition the central authority was ‘downwardly accountable’. Again, referring to the case, the creation and dissemination of program development and evaluation frameworks to schools was an example of the education department meeting its obligations to the schools, and indirectly to the people of Victoria.

Of course the exact formula by which the combination of strategies for achieving the balance between the information needs of auspicer and providers will depend on the nature of the system requiring evaluation. A key aspect of this new perspective on evaluation is a genuine commitment to evaluation findings being useful to providers in such systems. As one writer has noted

> demand for evaluation... may be motivated by a genuine desire to learn from experience, or merely to meet a bureaucratic requirement with no intention of using the gathered information. If the demand is genuine, evaluation information will potentially be used to help improve projects and programs....However, if information is merely being collected as a bureaucratic ritual, there is little reason to expect any quality in the data, any motivation on the part of those generating the data to ensure accuracy or any substantial trust in the enterprise (Rist 1995).

In short, providers need to see that investment in evaluation has a direct payoff for them and internal improvement focussed evaluation can deliver this payoff if it is well focussed. However, the realities of funding and resource allocation mean that system level authorities have a legitimate need for information which meets their information needs. This paper suggests some ways of meeting these accountability realities.

Finally, this brief discussion has set out the essential element of an accountability UP and DOWN framework. To flesh out the detail of this framework, a systematic critical analysis of the implementation of the School Accountability Framework in Victoria would be valuable in refining the theory presented here.
References


