Managing the Evaluation of Multi-level Multi-site Programs

John M Owen
Centre for Program Evaluation
The University of Melbourne

The Nature of Social Interventions

It is a truism that the design of an impact evaluation should take into account the nature of the evaluand or intervention under review. The proposed and actual nature of the intervention needs to be examined, not only so that we can be sure about how the program should and does work, but also to link processes to outcomes (Owen 2006). The importance of specifying a social or educational intervention contrasts with evaluation of medical and pharmaceutical interventions in which a drug or other treatment is simple and virtually a given.

Interest in the nature of social interventions has spawned what Shadish and his co-writers label as a theory of social programming. Issues that are of interest to evaluators within this theory include

> what are the important problems this program could address?, can the program be improved?, is it worth doing so? and if not, what is worth doing? (Shadish, Cook et al. 1991).

Shadish elaborates on this by suggesting that evaluators should be concerned with three elements
*internal program structure and functioning
*external constraints that shape and constrain programs, and

1

*how social change occurs, how programs change and how program change contributes to social change (p37).

I would like now to sketch key developments in the first of these elements. We have seen, over the past two decades, an increasing interest in program structure. The interest was originally spawned by the need to get over the 'black box' problem, that assumed that an intervention was in place when it was not. Such situations were common in the early years of large-scale school reform, where teachers simply did not implement the new curriculum, and so it had no chance of affecting desired student learning.

The extent of the problem was recognised by well-known theorists, and led to the development of techniques such as evaluabilty assessment. Working within a program management context, evaluability assessment was concerned with developing an initial assessment of the extent to which an outcomes evaluation was feasible. An important result of this concern was the development of process guidelines that led to a description of the program (Smith 1989).

Interest in specifying the nature of interventions moved on to take in what is now called program logic (or program theory), within the context of an emphasis on what I call process-outcome studies. These studies are motivated by the need to understand how a program operates in order to explain the pattern of outcomes. I have referred to such studies an important approach to impact evaluations (Owen 2006).

Clarificative Evaluation

A further development in evaluation practice has been an interest in the nature and portrayal of interventions in their own right, divorced from the link with outcomes, and effects of the implementation of the program on the host organisation (Rogers, Petrosino et al. 2000).

This interest falls under a form of evaluation that I have labelled Clarificative Evaluation. Studies under this heading are designed to assist program stakeholders' conceptualise interventions and improve their coherence, and so to increase the chance that implementation will lead to desired outcomes. The emergence of clarification in the

repertoire of evaluators is based on widespread experience that policy managers are faced by the following planning problems
*lack of time to fully research and articulate program plans
*a preference to base practice on implicit or tacit understandings
*a need to hurry implementation to meet political imperatives
*increasing complexity in the nature of social interventions
*multiple contributions of different agencies towards a given policy intervention, and
*increasing complexity of outcomes, for example to meet the triple bottom line of government expectations.

Single Level Programs

Clarification work of this nature almost always involves evaluators and program managers and deliverers working together. This is most easily undertaken on a single stand-alone intervention, which we call a small P program or project. Figure 1 illustrates the generic assumption of cause and effect for such a single intervention. In this Figure, S(I) signifies the outcomes of the program and S(F) are the (multiple) factors that affect these outcomes.

Figure 1. Causal Links in a One Level Program

In practice, program logic fleshes out these causal links to provide a program logic such as that contained in Figure 2. This example is a policy innovation designed to improve the productivity of organizations through the introduction of training programs for the workforce.

A feature of this logic is its linearity, which has both advantages and disadvantages from the point of view of program stakeholders and program evaluators.   Note that in this version, the strategy column has not been completed.

A linear portrayal provides a simple understandable version of how the program is meant to happen and the outcomes expected. The notion

of staged outcomes is made clear, which outcomes can be expected and when they will start to kick in. In addition, the logic links outcomes to an underlying rationale (theory), expressed in the form of program assumptions. These can be challenged at the logic development stage and forthwith. As Patton has noted, the logic enables sensible decision to be made about where and when to apply scarce evaluation resources, and provides a basis for planning a study in which causality can be incorporated (Patton 1997).

Figure 2. Workplace Skills Program Logic

WOPRKPLACE SKILLS TRAINING POLICY	PROGRAM LOGIC	
Policy		
objectives	assumptions	strategies
1. to create an awareness of and support for the creation of on-site structured skills amongst employers, unions and employees	1. employers need to know about workplace training in order to be committed to workplace training programs	
2. to delineate organisational and individual employer learning needs	2. workplace structured siklls training programs are most effective when they •involve management, unions and trainers in setting up processes, •are custom built following a careful skills audit and training needs analysis	
3. to train selected staff to lead structured skills training programs designed specifically for workplace contexts	3. structured skills training programs are likely to succeed if they are led by staff with an appreciation of the role of learning/training culture in workplace settings	
4. to provide programs that meet the work skill needs of employees in their specific workplace contexts	4. employee work skills are most effectively addressed via on-site structured skills programs tailored to their needs in the specific workplace setting	
5. to enhance employee knowledge and problem solving skills	5. knowledge and problem solving skills are an integral part of information used in the workplace	
6. to improve employee self confidence and morale	6. affective outcomes such as self confidence embeds learning gains in workplace performance	
7. to improve the productivity of an organization that support workplace training	7. modern organizations require flexible and educated workforces. These characteristics are essential for workplace efficiency.	
8. to maintain and improve national productivity across organisations	8. the wealth of a country is dependent on having a well educated and motivated workforce	

5

Multi-Level Programs

There is a temptation to oversimplify the nature of interventions, with the result that a linear portrayal does not adequately reflect the reality of the situation. The remainder of this paper is devoted to how I have been dealing with such situations as they relate to evaluation planning.

Let me reiterate two program related issues from earlier in the paper that are relevant here;
*increasing complexity in the nature of social interventions.
*multiple contributions of different agencies towards a given policy intervention, and
*increasing complexity of outcomes, for example to meet the triple bottom line of government expectations.

These are realities in the context of public sector interventions in Australasia today. The program logic in Figure 2 is presented in a linear diagram but in actuality the intervention described involves change strategies at two levels;
*that of a government department that is promoting the workplace skills policy, and
*at an individual agency level or site involved in the take-up of the policy. A more accurate version of the logic would incorporate these levels. These are formalised by the diagram in Figure 3.
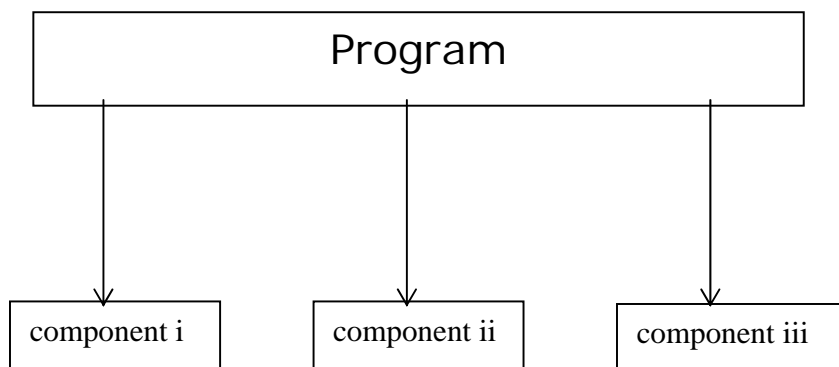
Figure 3. Multi-Level Nature of Programs
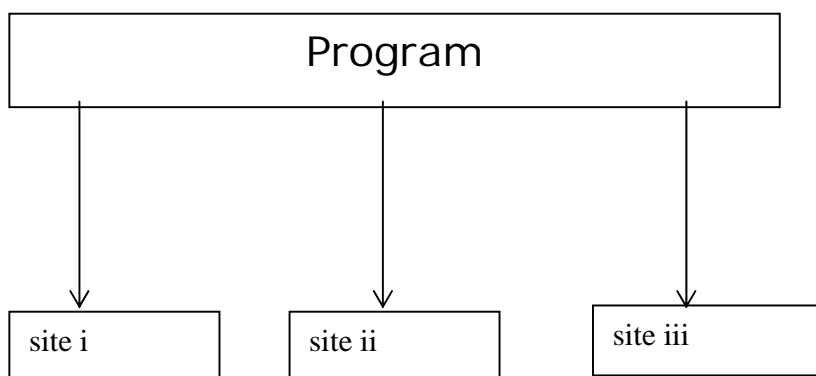
As indicated earlier, it is important to develop an evaluation design that acknowledges this complexity. This may involve a combination of common data from all sites in addition to more intensive studies of selected sites.

A second complexity involves programs that have distinct components, as represented in Figure 4.

Figure 4. Multi-Component Nature of Programs

```
                        ┌─────────────────────────────────────────┐
                        │              Program                     │
                        └─────────────────────────────────────────┘
                             │              │              │
                             ▼              ▼              ▼
                      ┌────────────┐ ┌─────────────┐ ┌──────────────┐
                      │ component i│ │component ii │ │component iii │
                      └────────────┘ └─────────────┘ └──────────────┘
```

Currently I am undertaking an evaluation of Orchestra Victoria's Rural Community Program (ROV). Using a range of data techniques, including interviews with orchestra management, we created separate program logics for three components of the program: (i) a free orchestra concert for the general public, (ii) a sponsored chamber orchestra concert for an identified charity, and (iii) workshops/ performance for talented school musicians. Providing the logic was of great interest to orchestra management in its own right, and has formed the basis for planning and implementation of the evaluation that is concluding about now. In the evaluation report, we present the findings that point to variations in the success of each component of ROV.

Multi-Level Programs and Contingency

A more complex multi-level multi-site intervention involves contingency. As some astute observers have noted, the achievement of desired outcomes at one level can be dependent on the extent of achievement, at an earlier point in time, of achievements at a higher level (Den Heyer 2001).

One can understand this with reference to the following diagram, which assumes three levels of influence of a given intervention (A, B and C). In Figure 5;

*I represents the impact of the intervention at a given level and F represents factors that affect that impact which can be attributed to that level. So, for example, A(I) are the intervention impacts at level A and A(F) are factors that influence these impacts at this level.

*the consequential level dependent time ordered aspect of the intervention is indicated by the relationship between impacts at one level and the outcomes at the higher level, for example the impact at level B is dependent on the prior outcomes at level A

*in addition, level B impacts are also dependent on factors which can be attributed to B level factors

and so on.

Figure 5. Framework for time dependent and hierarchical levels of influence of a social intervention
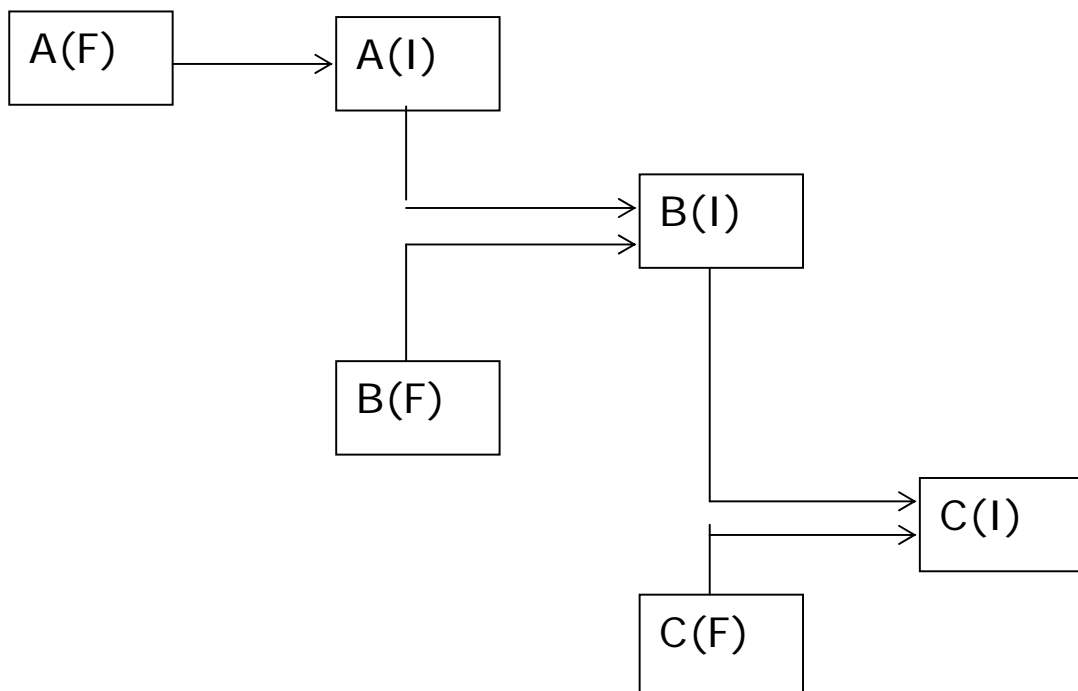
A variation on this paradigm is that, for a given A, there could be multiple Bs and for a given B, there could be multiple Cs, and so on.

During the past three years, I have served on the evaluation committee for the Mind Matters national curriculum project. Mind Matters is designed to encourage schools to adopt and implement teaching and support programs about mental health issues. The intervention is located at the Australian Principals Associations Professional Development Council (APAPDC), in Adelaide.

The intervention involves a change strategy designed to influence schools to take up Mind Matters at three levels; the
*system level, through the incorporation of Mind Matters principles in education department frameworks, with impact indicators such as the number of schools adopting the MM ideas and materials. This is the A level in Figure 5.
*school level, through the incorporation of MindMatters into the whole school curriculum, with impact indicators such as a commitment into school policy or strategic planning (level B).
*classroom level, the teaching about Mind Matters in teaching/learning settings, with impact indicators such as the effects of the teaching on the knowledge, skills and attitudes of students (level C)

Using Multi-level Multi-site Framework in Evaluation Planning

The MM evaluation committee was charged with commissioning evaluation work that would provide an impact of MM as it was rolled out over the past three years. The committee commissioned the following studies
*system level:
        (i) take-up of MM in schools across the country
*school level
        (ii) strategies used to embed MM in the school curriculum, and
        (iii) ways in which schools dealt with students of high need, through an extension of MM (known as MM+)
*classroom level
        (iv) teaching of MM resource units (eg Understanding Mental Illness) and their effects on students.

One can draw the links between these evaluation strategies and the diagram in Figure 5. A vital aspect of this strategy was that the findings provided complementary findings about impact. At the end of the evaluation period (July 2006) MM stakeholders were provided with a comprehensive set of findings about MM. The evaluation committee also performed an educative role to some of the represented

9

AES Head Office: PO Box 5223 Lyneham ACT 2602 **ABN** 13 886 280 969
Ph: +61 2 6262 9093 **Fax:** +61 2 6262 9095
Email: aes@aes.asn.au **Website:** www.aes.asn.au

stakeholders, who had limited experience of the complexities surrounding the introduction of new ideas into schools. There was a need for lessons about the extent to which causality links could be claimed. For example, there was a tendency for some stakeholders to argue for changes in student knowledge about mental health as a system level outcome, despite the fact that take-up across the country was patchy. We took seriously the need for contingent causality. For example, the study, which examined the embedding of MM in the school curriculum, was able to draw defensible links to the extent of professional development provided to teachers by MM.

An additional important point is that each of these studies involved different methodologies and data analytical techniques. For example the national survey used paper and online surveys, the embedded study was naturalistic, using observation and interviews in schools over time, and the classroom study relied on teaching-learning situation observations and student tests. This serves to emphasise my strong and abiding concern that methodology should be selected to answer the evaluation issues, not vice versa.

Conclusion

You may be able to discern a pattern in the discussion, which is summarised in the diagrams provided in this paper. Figures 1 and 2 describe first a general schema and then the program logic for a single level intervention. This was expanded to a multi-level intervention schema in Figures 3 and 4, from which program logic statements were developed for Orchestra Victoria. Figure 5 expands the idea to a contingent multi-level logic. Alas, in the case of the relevant case, the MindMatters example, a logic has not been developed. I have to admit that the evaluation committee of MindMatters used an implicit logic to commission the evaluation studies described here.

I began this paper with a wish to extend my thinking about logic statements that can guide the implementation of complex policy and program innovations. A second and often related issue is the planning and implementation of effective evaluations of these interventions. It would be useful to continue thinking along these lines. This would be a service to those responsible for managing the change processes involved in dissemination, in addition to assisting those responsible for planning and conduct of studies that respond to the information needs of policy and program stakeholders.

References

Den Heyer, M. (2001). The Temporal Logic Model. Ottowa, Canada, International Development Research Centre.

Owen, J. M. (2006). Program Evaluation: Forms and Approaches. Sydney, Allen and Unwin.

Patton, M. F. (1997). "Developing Program Theory." Evaluation News and Comment **6**(1): 18-20.

Rogers, P. J., A. Petrosino, et al. (2000). "Program Theory Evaluation:Practice, Promise, and Problems." New Directions for Evaluation **87**: 5-13.

Shadish, W. R., T. D. Cook, et al. (1991). Foundations of Program Evaluation. Newbury Park, Sage.

Smith, M. A. (1989). Evaluability Assessment of the Maryland Water Quality/Quantity Program, Personal communication from the author.

dar fin 090806.doc