Evaluation journal of Australasia

Volume 8 Number 2 2008
Editors: Rosalind Hurworth/Delwyn Goodrick

The *Evaluation Journal of Australasia* is published twice a year and is the official journal of the

Australasian Evaluation Society (Inc.)
PO Box 5223, Lyneham ACT 2602
Australia
Tel: +61 2 6262 9093
Fax: +61 2 6262 9095
Email: aes@aes.asn.au
Website: www.aes.asn.au

Intending contributors should consult the 'Notes to Contributors' (p. 60).

AES membership form (with current fees): www.aes.asn.au/membership

Library subscriptions: A$100 plus $10 GST for Australian libraries

Copy editing:
Helen Andersen (Style Matters)
Cover, page design and layout:
Studio 673 P/L
Printing: Union Offset Printers (ACT)

ISSN 1035-719X

AUSTRALASIAN EVALUATION SOCIETY INC.

Contents

EDITORIAL

Once again we are able to bring you a diverse range of material within the covers of this issue and viewpoints derive from both international and local authors.

So, to begin we are very fortunate to have an article by renowned Ray Pawson (of Pawson & Tilley, *Realistic Evaluation*, fame). In this article he delves into the 'black box' of evaluation and urges the rest of the evaluation profession to follow. He calls on us to examine the presence of frequently overlooked program mechanisms, 'so deeply buried that they are almost invisible'—because they are tacit, and usually taken-for-granted. Yet these mechanisms are often responsible for the impact of interventions. The purpose of the article is to present strategies for their investigation and to think about how to make the tacit explicit, especially for program design and policymaking.

This lead article is followed by one of the keynote addresses at the AES International Conference held in Perth last September. Presented by Scott Bayley, it was particularly well received. Many people asked for it to be printed, not least because of the growing interest in the setting for the paper, that is, international development. By answering a series of questions, the article looks at how to improve project, program and policy performance in developing countries. Responses are drawn largely from the experience of the Operations Evaluation Department of the Asian Development Bank where Scott works as an evaluation specialist.

Coming nearer to home, but still thinking of remote and challenging communities, the next article considers whether participatory approaches in evaluation can be applied successfully in more difficult circumstances. Catherine Spooner and her colleagues describe a case study of a crime prevention program for rural Aboriginal boys and, in doing so, examine the factors that seem to contribute to successful participation.

Also considering challenges is Anne Markewicz who writes about issues for the evaluator when dealing with political and policy contexts while still thinking of the stakeholder's role. She also suggests that there has to be some reconciliation between the influence of the participating key stakeholders and the credibility of the evaluation process.

Then we continue the series concerning useful resources for evaluators. This time these are couched in an attempt to unravel terms associated with clarifying programs. Therefore, the preamble considers how terms such as 'evaluability assessment', 'program theory' and 'program logic' have developed before providing pertinent reading. In addition, rather than give a long list of readings purely in alphabetical order, material has been grouped by discipline to assist evaluators working in different sectors. We hope now that readers will send in other topics that they would like to see as a basis of a resource list or send in their own lists.

Finally, there are half a dozen book reviews. We have received several more but due to lack of space, some of these have been reserved for later issues.

We now await your submissions for issues in 2009. Many excellent papers were presented at the Perth conference which would make interesting reading. So please consider sending us a contribution.

Rosalind Hurworth and Delwyn Goodrick
Editors

Invited paper

Invisible mechanisms

Ray Pawson

Ray Pawson is Professor of Social Research Methodology, University of Leeds, England

The mandate for evaluators to 'look inside the black box' of an intervention has become a familiar *and* heeded cry. So whether it is process evaluators with their logic maps, or theories-of-change researchers with their intervention stepping stones, or realists with their context, mechanism outcome configurations, searchlights have been aimed into the gloom. So much so that a contemporary sounding riposte, 'been in there, done for that' might be deemed to reflect the current state of play in evaluation research.

This article, nonetheless, warns against complacency. It peers even further into the darkest reaches of the inky blackness, and urges other researchers to follow. It throws light on the presence of an overlooked set of program mechanisms, so deeply buried that they are almost invisible. The processes I have in mind are missed because they are tacit, mundane, over-familiar, and taken for granted and, as result, they often overlooked. And yet they are often responsible for a goodly part of impact of a goodly number of interventions.

As such, they deserve a sustained program of research and this article sets out a brief agenda for such inquiries.

Latent procedures in implementation and evaluation

Policymaking is energised by the hot new idea. Attention is thus drawn immediately to the unique properties and powers of the new 'measure', 'treatment', 'therapy', 'mechanism of action', or 'theory of change'. To be sure, other eyes are always on the prize, namely impact on the intended outcome. Accordingly, interventions find support and are brought to life if there are persuasive reasons to believe that a new-fangled idea might have a significant leverage on a long-standing problem.

But what happens next? The machine takes over. The intervention and its evaluation are assembled in a series of standard procedures. The program has to be organised and delivered—sites are mulled over and selected, resources and staff roles are allocated, participants are recruited and processed. The evaluation, too, has a rhythm—cases are sampled, baseline parameters are assembled, processes are inspected, outcomes are assessed. The working hypothesis here is that these routine features, the generics of programming and evaluation, often have as profound an influence on program participants as do the big ideas.

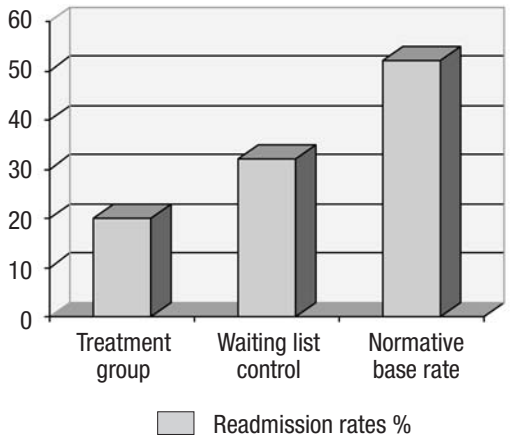
People enter programs at the margins and sometimes quite tangentially: they have life outside programs; there are always other programs and life offers many new opportunities besides programs. And once within the ambit of a program there are many opportunities to quit or stay, and even within the camp followers, there is a range of commitments from 'passing interest' to 'abiding passion'. There are many such collateral pathways for so-called 'program subjects' to consider, and the mechanics of this movement from marginality to provisional membership to full membership has been overlooked in evaluation research.

Let us begin with a Cook's tour (i.e. swift but wide-ranging) around these subdued, contemplative quarters. *Where and when do invisible mechanisms begin to reveal their inscrutable faces?*

Behind bars

Long ago, whilst still sporting my evaluation 'L plates', I came across a simple chart that caused me to puzzle over it. It is reproduced as Figure 1. It shows results from a pilot investigation of a so-called 'cognitive skills' program aimed a reducing recidivism in a group of inmates in Canadian federal prisons (Porporino & Robinson 1995). Even within this population one cannot require subjects to attend programs. Subjects are volunteers and the conventional way of evaluating impact whilst countering the associated 'self-selection' effect is to run a trial comparing treated subjects with those who have also volunteered but are kept, often surreptitiously, on a 'waiting list'.

FIGURE 1: FINDINGS OF A PILOT INVESTIGATION OF A COGNITIVE SKILLS PROGRAM



The conventional reading of these findings is that the treatment indeed offers benefits, as can be seen from the lower rates of return to prison of the experimentees (20 per cent), as opposed to the waiting-list controls (32 per cent). The fact that we are dealing with a reasonably well-disposed group is also demonstrated by the gains in the untreated volunteers as opposed to the baseline rate, where the revolving door of reincarceration revolves at over 50 per cent.

This interpretation of the data has basically two ingredients: the treatment with its power to change and the given predispositions of volunteer subjects. But neither of these factors explains fully what is going in the control group. What are the inmates doing and thinking whilst waiting? Are they growing impatient or are they learning forbearance? Is anger mounting? What's the point?'

The answer, of course, is that we do not know. As far as I'm aware no serious research attention has been directed at the folks on hold.[1] But what should, at the very least, be contemplated are the powerful and potentially life-changing emotions represented by the participants' responses to the program. And as such, they represent our first sighting of invisible mechanisms.

I am trying to suggest, via the inmates' notional thoughts, that the members of the 'control group' are hardly in the state of repose that is suggested by that term. They are in the side-wash of the intervention and this location in itself can trigger impatience or anger or forbearance or fortitude. Lack of support does not turn them significantly back to the pack. At the very margins of a program, we witness a complex calculation. A future course of action is contemplated, an opportunity along the way is spotted and then stymied, and other ways of continuing in the direction of travel are apparently discovered. Here then is a brief glimpse of the latent action of a program—if all this reckoning and re-reckoning happens to subjects supposedly twiddling thumbs in the control condition, what other thought permutations transpire over the life-course of an intervention?

We stay within prison walls for my next illustration. The theme, however, continues—programs are not things, not dosages but complex social situations opening up a potential menu of choices within choices. What else might entice the subject to take a seat? What other imperceptible offerings might be served up? Duguid (2000), a practitioner, researcher and theorist of 'prisoner education', comes up with the following counter-intuitive recipe: offer education as a rehabilitative therapy and it will fail; offer education for its own sake and it will carry over into rehabilitation.

The idea was made manifest in the Simon Fraser Prison Education Programme. This campus-in-the-prison program ran for two decades in several Canadian penitentiaries. It was as avowedly 'mainstream' as far as the difficult conditions of imprisonment allowed. On offer were degrees and diplomas from Simon Fraser University; faculty crossed between university and prison; provision was full-time and year-round; the education block was separate and for the most part self-policed; credits were transferable to outside institutions; prisoner 'graduates' could become course tutors. And perhaps above all, talk of therapy, counselling, reform and rehabilitation was 'off-limits'.

How could these bread-and-butter features of campus life influence rehabilitation? In the following

passage, Duguid (2000, p. 230) provides us with a hearty list of invisible mechanisms that may account for program success:

- Community, self and authenticity: An ethical stance towards the prisoner based on interacting with him or her as a subject rather than an object. In its structural form this often centres on creating a democratic participatory environment within which the program operates.
- Bonds with the conventional world: A politics of prison programming that stresses the connection between the specific initiative and an institutional affiliation external to the prison and the criminal justice system.
- A structural approach that relies on diversity and complexity rather than singularity and simplicity, acknowledging that prisoner needs are many and unique and the intervenor's skills are various and limited.

It is perhaps useful once again to make the primary proposition clear. Neither Duguid nor I dispute the idea that other and perhaps more tangible aspects of prisoner reform, such as the improvements in employability or cognitive ability or social skills, are brought into play via such a program. The contention is that these gains are facilitated because of a culture, and it is this unadulterated, enduring and authentic taste of a learning environment that also offers the inmates a second chance.

Can such a claim be demonstrated?

The issue of separating and attributing differential causal powers to the multiple mechanisms embedded in a program represents the toughest challenge for evaluation methodology. The third section of this article returns to this issue more generally. For now, we note the solution lies in the comparison of the differential success of different pathways through the program. Duguid's team (2000) used this approach to isolate the varied success of different subgroups and of different encounters with the intervention (following the realist dictum, 'what works for whom in what circumstances and in what respects'). But I am less sure how successful the research was on isolating the impact of the hidden mechanisms.[2]

Perhaps the nearest portion of the inquiry concerned the fortunes of the 'theatre group'. The specific argument is that such an environment embodies Duguid's latent forces (above) in a concentrated form, namely: in-depth emotional and intellectual study required to play a role; the need for spontaneous civility in mounting a production; a public sphere in which collaborative decision-making was of the essence; sustained detachment from the institutional imperatives (of both control and rehabilitation) within the prison. A group of prisoners rated as decidedly 'high-risk' prospered well on this particular program. Their predicted rate of return to incarceration was a substantial 64 per cent compared to the eventuality at 45 per cent (Duguid 2000, p. 245). And whilst there was a generally improving trend throughout the program, these 'hard cases' thrived significantly better here than on other schemes.

To be sure, this is a mere fragment of evidence but it is a sign of latent mechanisms working their way through to manifest outcomes. If, moreover, one examines the direct testimony of the prisoner students, the action of the non-remedial remedy shouts out loud. For this, I return to some interviews of my own (reported in Pawson 1996) conducted in a UK prison, but on the same topic of how education might promote change in inmates. These were hardly cosy chats. Indeed the exchanges were somewhat fiery as the men clarified the matter of who was controlling whom. The common theme was about how education itself hadn't rehabilitated them but rather fired an interest already inside them, about how it deepened a process of self-scrutiny that was already underway.

... On the therapist's couch

For our next case study, we turn from the province of 'Nothing Works' to the domain of the 'Dodo's Verdict'. Alice, during her adventures in Wonderland, comes across a curious competition officiated by a dodo bird. It is a simple enough contest, a race around a lake. The twist is that no one bothers to measure times, distances, placements, and so on. Instead, the dodo opines: 'Everybody has won and all must have prizes'.

This same unflattering verdict has been bestowed on psychotherapy. There are many, many different therapeutic schools. One count, made 40 years ago (Parloff 1986), estimated the number at 418. A thunderous and longstanding critique argues that the specific techniques associated with specific schools (e.g. Freudian, Jungian, Rogerian, Adlerian, behavioural, cognitive, gestalt, existential, etc.) serve very limited purpose and that most of the positive effect is gained due to therapeutic *relationship*.

This hypothesis known as 'common factor theory' associates positive change with 'non-specifics' emanating from purposeful, warm, respectful, tailor-made, one-to-one relationships between practitioner and client.

Psychotherapy has always been a hot topic for evaluation. As related studies grew, they evolved a comparative component in which the efficacy of rival or alternative treatments was investigated.[3] And as these 'x versus y' studies gathered pace it became possible to conduct meta-analysis of 'comparative treatment studies'. Studies of this ilk by Luborsky and colleagues in 1975, and repeated in 2002 with a much larger sample of primary studies, came down heavily on the side of the dodo bird verdict. Very few primary studies demonstrated the superiority of one treatment over another. In the round, the meta-analysis estimates that the effect size attributable to specific therapy techniques weighs in with a Cohen's *d* coefficient of only 0.2

(small and insignificant in lay parlance). By this interpretation all therapies are more or less equal and almost all should win prizes (or perhaps be amalgamated, simplified and demystified!).

It should be said at once that this interpretation is not without problems. Meta-analysis works at a high level of aggregation (Pawson 2006, Chapter 3), tending to pool a diverse medley of program characteristics in order to calculate a mean effect. Chambliss (2002) picks up this criticism, pointing to the dangers of examining 'average differences between all sorts of treatments for all sorts of problems'. It remains possible that specific techniques may well be shown to be effective—if examined in specific respects by specific measures, and if one differentiates particular subgroups of subjects in particular circumstances. There is no need for us to referee this particular dispute. Whilst debate continues to rumble about the precise arithmetical contribution of the common factors in common factor theory, no one denies the import of the 'non-specifics'.

It is much more useful for present purposes to look at attempts to discern and itemise the collective content. A number of researchers have attempted to locate precisely what is common in common factor theory. Following an extensive review of the client's experience, Tallman and Bohart (1999, p. 106) sum up that:

> from the client's perspective, the most important aspects of therapy are the non-specific factors—the personality of the therapist; having a time and place to talk; having someone to care, listen and understand; having someone provide encouragement and advice; having someone to help you understand your problems.

What catches my eye about this list of invisible mechanisms is just how many of the features are also claimed in programs well removed from the formal therapeutic encounter. They are part of the rationale for peer education for drug abuse, buddy programs in prison, mentoring schemes for disaffected youth; coaching projects for would-be women executives, and so on.

Strupp (1986), from an earlier generation of research into 'non-specific factors', pinpoints a different causal mechanism, namely the guiding force of 'theory' in galvanising the psychotherapeutic process. The argument here is not about the express thoughts any of the 418 (or more) schools of thought. Rather it focuses on the very existence of 'a theory' in underpinning and guiding the intervention. The theory provides therapists with a resource capable of organising and planning the treatment. It also provides them with intensity and depth of purpose to keep them engaged over a long time. Finally, the theory may be said to certify and legitimate their particular approach. These capacities are no small matters from the point of view of the client.

Generally speaking, people who volunteer for programs are seeking to 'work out an understanding'. They are always on the look out for theories—the operative issue being that it is not only therapists who trade in explanations. When it comes to implementing programs—cops have theories, teachers have theories, big brothers have theories, ward sisters have theories, and safety inspectors have theories. All of them may gather adherents.

We are not yet done with the non-specific effects, for there is another important conjecture claiming that they kick in before treatment even begins. Rather than garnering lessons only from academic research, let us once again hear it from the horse's mouth. In this case practitioner wisdom emanates not so much from the couch but the floor mat. Latey is the author of *Muscular Manifesto* (1979) and his thing is 'movement therapy'. Whilst specialising in so-called 'bodywork', he has clearly spent time reading the minds of his clients. In the following passage, Latey (2001, p. 149) goes well beyond the customary recognition that patient 'self-motivation' is crucial by going on to describe how this vital spark might be encouraged and enhanced by creating certain preconditions to receiving treatment.

> I believe it helps if patients have had to surmount some difficulties in order to get to see the practitioner, as follows:
>
> - a wait for an appointment at a time that may not be easy for them
> - some directions to follow if practitioners are off the map of their usual movements
> - the effort of organising their account of the problem
> - preparing to be questioned, examined and treated in the first session.
>
> The fact that they are willing to pay for treatment, however small the fee, makes a considerable difference … people expect to pay and do not count the cost when their health is at stake. Timeliness is also crucial. So it helps if patients have understood the problem is not going to clear up by itself, and they have reached a point where it must simply be sorted out. All the better if they have also abandoned previous attempts at treatment with enough time for it to be obvious that they have failed.'

Here then is another raft of invisible mechanisms that may contribute crucially to program efficacy. We have already encountered perverse effects within waiting list controls and, interestingly, Latey (2001) also recommends being positively artful with the 'keep 'em waiting' rule. However, this is but one of a number of other *pre-intervention* strategies that may be transferable—intensive openings, speed off the mark, quick wins, immediate active role for client, arrival in the last chance saloon, and not forgetting usage of that old advertising slogan, 'You've tried the rest now buy the best'.

... In Fagin's Den

The roll call of invisible mechanisms continues with an article by Smith, Clarke and Pease (2002) under the title, 'Anticipatory benefits in crime prevention'. This is a useful extension to our thinking about tacit powers of interventions because the goal in this domain is to control and constrain potential action. Our previous examples cover attempts to facilitate fresh thinking and behaviour. Invisible mechanisms can operate in both directions.

The authors commence their case with the stunning quotation (used as the epigraph to this article) on the seemingly potent program theories of yore. Their central argument is that we should never jump to conclusions about the 'self-evident' causal powers of interventions. Crime reduction, for the most part, works by persuading potential offenders that the risk of apprehension and arrest increases under a newly installed program. Perception is the key and thus it may be that the *threat* of action of an intervention is as powerful of as the *specifics* of action.

Many programs appear to show improvement (crime reduction) before the program is up and running. Indeed, some seem to work without them being enacted properly. This hypothesis is examined on the basis of a review of the crime prevention literature. A search was undertaken locating studies that contained time-series data sufficiently powerful to distinguish crime fluctuations before, during and after the introduction of prevention programs. Fifty-two such reports were uncovered that revealed an unexpected pre-initiative drop in crime statistics. Of these 22 had strong prima facie evidence that allowed causal attribution to 'something' occurring within the early inception of the scheme.

For instance, a study of the effects of security cycle patrols on parking lot crime showed that announcing the scheme was followed by a reduction in crime. Ending the scheme, moreover, did not result in an immediate increase in crime (Barclay et al. 1997). Further examples relating to the pre-installation of CCTV cameras, security devices, alcohol testing, physical layout improvements and so on are assessed and corroborated in the review (Smith, Clarke & Pease 2002, pp. 75–76).

So what is the 'something' that could account for these unanticipated anticipatory effects? Smith and colleagues (pp. 78–79) list 10 possible mechanisms, which I further summarise thus:

- *Evaluation artefacts*. These include some time-honoured measurement headaches such as regression to the mean, difficulties with calculating moving-averages, and the perennial problem of seasonal shifts in recorded levels of crime.
- *Practitioner and subject effects*. These include improvement in knowledge and motivation of the local population and police officers on the announcement of a new scheme, which translate into increased determination, greater diligence and better performance in advance of the initiative.
- *Offender effects*. These include both the 'over-anticipation' effect, in which equipment is supposed to be operational before it actually is and the 'disinformation' effect in which publicity and hearsay carry the impression that a powerful, covert program is already in place.

Again we see a catch-all description, in this instance 'anticipatory effects', netting a miscellany of possible mechanisms. The point for reinforcement is that none of the above are part of the intended measure; all are part of the implementation and evaluation apparatus—and thus all are open to further and more mindful manipulation by program planners.

And it is this respect that the latter item in their list excites the attention of Smith and colleagues. If we think of crime, at least some crime, as an 'intelligence-led' operation then 'counter-intelligence' becomes on option in its curtailment. There is probably an element of this idea in all policing. Smith, Clarke and Pease (2002) consider the example of so-called 'informants'. These supposedly lurk in the underworld telling the police what putative offenders are up to. Just as beneficial to the strategy of risk enhancement is for them to inform putative offenders what the police wish them to think is happening.

How could such a mechanism be embodied in a formal intervention? The active ingredient in all the cases reviewed appears to centre on the circulation of information—getting the word out on the street. The optimal working example is probably the action of 'decoy vehicles' in reducing car theft. Cars and vans, similar to those favoured by thieves, are parked in high vehicle crime locations. They are fitted with technical devices making it possible to track or, sometimes, trap the intruder. Whilst this immediate and tangible mechanism is what does the job in apprehending offenders, it seems that hearsay buttressed by media campaigns is the invisible, diffusive mechanism that really brings down overall rates in a locality (Sallybanks 2001). The scheme makes would-be offenders ponder precisely at the point when they normally sense an easy picking. And that rumination is deepened if they have in mind television pictures of the speculator and embarrassing failure associated with being so outwitted.

... In swallowing the pill

Our search for latent effects ends with a critical case, namely clinical interventions. In the orthodox medical model, the causal powers of the treatment reside at the physiological level, allowing medication to attack viruses, kill cancerous cells, relax blood vessels, heal bacterial infections, boost the immune system, and so on and so forth. Abutted to this viewpoint, somewhat uneasily, is the complementary perspective arguing that much that is efficacious about treatment lies before, during and after the

swallowing of the pill. We thus end our tour in these disputed waters and with the most famous invisible mechanism, namely the 'placebo effect' in medical trials. Despite Moerman's ironic quip (2002, p. xiii) about how easy it would be to write a placebo book—'because it would have nothing in it'—it turns out that there is a massive literature on the said topic.

It is useful to begin by reprising a note made earlier about life on the margins of a program trial—in this case as a member of a control or untreated group. Time does not cease for such assemblages. This is how Moerman (2002, p. 26) goes on to argue that it is 'logically and conceptually impossible to have a no-treatment group' in which disease runs its natural course[4]:

> In order to do a trial, people have to be recruited and diagnosed for the condition under study: they receive some sort of examination, maybe an intensive and dramatic one. They give informed consent, perhaps after reading a long and complex document describing the study, the various treatments under review, and so on. They are then randomly assigned to three conditions: drug treatment, placebo treatment, of no treatment. It's not clear what one will tell the group getting 'no treatment'. Certainly their participation can't be blind to them: they know they aren't getting any drugs or placebos; a reasonable inference might be they are healthy enough not to need any. And after that there has to be a follow up, an assessment of the condition of the subjects after some period of time, or a diary of symptoms has to be kept. While these people have not had pills, they have had a great deal more than nothing.

Indeed, this quotation opens the door to even more possibilities. It describes the 'no-treatment' control group, but it is obvious that the 'placebo' control group has similar levels of contact with the experiment, if a slightly different conundrum to decipher about their place within it. Moreover, it is at least a possibility that dearth of treatment in the absolute null condition could promote despair about absence of hope rather than optimism about marginality of need. The sensible inference to draw from the above is that the treatment process is a long and complex business capable of attracting diverse inferences in the minds of inference-making subjects.

This proposition provides the theme for this section, for the history of placebo research is a story of how the core idea of 'placebo' has, by dint of close empirical research, been broken down into a number of component social and psychological process.[5] What we learn about these 'meaning effects', operating in clinical conditions where they are often considered marginal and a nuisance to boot, carries important lessons for social programs, where they are much closer to centre stage.

For empirical backing here I rely on several research reviews of 'treatment dynamics', 'doctor-patient communication', 'self-healing', and so on (Moerman 2002; Stewart 1995; van Dulman & Bensing 2002; Vermeire, Hearnshaw & van Royen 2001). I draw rapidly and selectively from these in identifying a handful of tacit encounters that might be particularly significant. As the case for invisible mechanisms consolidates, the material uncovered in the clinic has strong echoes of latent processes already described.

It is by now a common cry that greater patient involvement in their own treatment may lead to improved outcomes. Stewart's (1995, p. 1422) discussion of experiments on patient choice reveals some subtle distinctions that help to clarify the mechanism involved:

> In one study (Morris & Royal, 1988) the fact that a woman was able to choose the kind of breast surgery to have [mastectomy or lumpectomy] was not found to be related to emotional health outcomes. In another (Fallowfield et al., 1990), going to a surgeon who permitted but did not force the choice, *was* found to be related to positive outcomes. I would suggest, therefore, that it was not simply the decision making power of the patient that was effective but, rather, the provision of a caring, respectful and empowering context in which a woman was enabled to make an important decision with both support and comfort. (References not claimed for this paper.)

The dilemma portrayed here, whilst horribly specific, harbours a striking resemblance to the lot of many program recipients. Subjects always choose but rarely choose the choices open to them, or know that much about them. Conversely, the deeper the contemplation of the choice, the more informed the choice, the more determined is the subsequent pursuit of the choice. It is no accident, for instance, that this process echoes earlier reflection on collaborative decision-making on offer in some prisoner education programs.

Choices cut both ways of course and another literature on (lack of) compliance with treatment can help us build our model of program pathways. Lack of adherence to treatment is another bugbear of the RCT. Dracup and Meleis (1982) conducted a pioneering inquiry attempting to fathom reasons for an initial successful and subsequent unsuccessful trial of the same drug for reducing blood pressure. It turned out the hypertension regimen was followed by 80 per cent of patients in the former trial compared to only 50 per cent in the latter. Since these early studies, research has turned to the reasons for non-compliance and that rationale has been explained in a theory known as the 'health belief model' (McGavock 1996). When confronted with an illness, people try to deal with it through their own experience, resources and folk wisdom. Later, when they arrive at a consultation, patients still bring with them a set of ideas and expectations about health and illness. There is no fixed point at which lay knowledge concedes to professional expertise, with the result that adherence

to treatment can wobble throughout the treatment. Donovan (1995), for instance, reports on patients' self-experimentation in modifying the prescribed drug intake to diminish the risk of side effects and in order to discover the lowest drug dosage that seems effective for them.

As long as there are doctors, their convictions will play an active role that cannot but influence medical outcomes. Moerman (2002, p. 45) provides an interesting example of the doctor-as-an-active-ingredient, which provides an explanatory glimpse into its operation. One routine context that prompts the 'physician effect' is the constant throughput of new drugs and treatments. These, of course, excite the interest and expand the knowledge of the physician. So much so that there is often a bizarre regression in which old drugs seemingly become less efficacious as new ones come along. Moerman reports on trials of drugs for ulcers. The original trials on the first drug (Tagamet) resulted in 72 per cent of patients being healed. Seven years later a new drug (Zantac) came and its trials showed a slightly improved 75 per cent rate of healing. A contemporaneous, second-wave of trials on Tagamet were also performed and, curiously, the efficacy of the same drug across the same population had dropped to 64 per cent. Enthusiasm, as they say, radiates and it seems that interest in the new drug was balanced by a disparaging of the old. These drugs are dispensed by injection and tested by endoscopic examination and there are opportunities at both ends for the subtle transfer of anticipation. The offer of a 'new and improved' regime may well be contagious.

Again, I charge Moerman (2002, p. 45) with providing us with the lasting lesson: 'Doctors know lots of things. Many of the things they know they are unaware of knowing (as is true for many of us in this life). But it is the depth of their conviction that conveys to patients power of their treatments'.

Whilst clinical treatment is vastly different from other forms of social interventions, there is commonality across the four case studies examined here. What the placebo paradox and the ensuing inquiries tell us is that the path from illness to cure, should it materialise, is a journey rather than a turning point. A whole range of collateral—one might indeed say complementary—mechanisms facilitate the journey. Best practice in medicine rests on biological and physiological change but, as elsewhere, it also involves interweaving an array of psychological and social processes—some of them rather more opaque than others.

Unspoken mechanisms articulated

This section attempts to produce an abstract model of the pathway of change describing the cumulative, progressive, iterative transformations that typify the vast majority of social interventions. I am already blue in the face with arguing that social programs do not work through Pauline conversions, divine deliverance, instant redemption or miracle cures. They work by persuading subjects to change. And subjects, from the very beginning, will be relatively recalcitrant or willing. Subjects on the threshold of a program will ponder, wait, figure, investigate, and change their minds. Subjects over the threshold will dive in, tread warily, pull out, dawdle, support, sabotage, take over, malinger, proselytise, and so on. Programs work to the extent that they can shift the tide, moving sufficient numbers of the marginal and refractory into compliance and commitment with the intervention goals.

Figure 2 attempts to map the pathway of subjects from first contact with a program to exit. It assumes that a long journey is involved and that subjects will fall by the wayside in many an intervention. It contains eight staging posts, reflecting changes as subjects move from marginality to membership. Each stage is depicted as a decision point for the program participant and at each stage includes the program processes designed to propel the subject onwards.

As such, it is a distillation of the many, many processes and preconditions captured in the previous sections reduced to a single sequence within an imaginary or 'ideal-type' program. The model is thus a middle-range theory (Merton 1967; Pawson 2000) in the classic sense **in** that it seeks to confederate a range of distinct empirical uniformities into an abstract model. The model allows us a clearer view of the underlying process, which can be formalised further by producing auxiliary hypotheses about how the sequences intermesh and what will happen if they do not. These propositions then provide an explanatory ensemble for predicting and planning other implementation pathways in other programs.

More concretely, one can say that the model already embodies the actions of a prisoner wondering how to go straight, an osteopath seeking to drum up more business, a cop spreading intelligence about a new crime reduction gizmo, a patient thinking about whether to tinker with treatment, and so on. All of these activities and more are captured by the abstract formulation of the model. And when the continuity of underlying constituent processes is appreciated more fully, the expectation is that the model will be transferable. It may be able to teach us something about the choices of program stakeholders thus far unconsidered, such as surgeons defending individualised waiting lists, disengaged youth wondering whether to bother with the latest government training scheme, company schemes trying to attract more minority applicants, and so on.

The top row depicts the decision points through which contemplative subjects pass. At all stages subjects are choice makers. The flow through the program may be continuous and to plan. Or, it might stall, short-circuit, or backfire. Or, to coin a phrase, it may move two steps forward and then one backwards. The bottom row reflects upon the opportunity for program planners and practitioners to encourage and propel each choice in the right direction. In realist parlance, this lower

FIGURE 2: PROGRAM PATHWAYS—WEAVING THE SUBJECT INTO THE PROGRAM

	1	2	3	4	5	6	7	8
Decision pathway for program subject	**Awareness**	**Horizon scanning**	**Path selection**	**Gaining access**	**Initial reflection**	**Refraction/ compliance**	**Full membership**	**'Graduate' and exit**
	Subject becomes conscious that they might have a 'problem'	Subject casts around for solution for problem	Subject accepts 'diagnosis' located in specific program	Subject decides to give program an initial try	Subject reflects on first experience of program	Subject questions the worth of staying in the course	Subject buy-in and commitment to goals of the program	Subjects enact program goals and leave carrying durable lessons
Facilitation opportunity for policymakers and practitioners	**Flag-raising**	**Contrastive publicity**	**Red carpet**	**Pull rug**	**Quick wins**	**Continuity**	**Empowerment**	**Attest, distance and recycle**
	General publicity to name problem, climates of awareness, moral panics, buzz wordage	Promote intervention theory 'You've tried the rest now choose the best' Success of 'people like you'	Access and recruitment organised Intervention modalities explained and potential gains clarified through exemplars	Emphasise subject responsibility Test patience, commitment and effort	Provide access to expertise and interpersonal contacts, offering short-term personal gains, not necessarily related to final outcome	Stress participatory responsibilities Retest commitment and effort	Cede elements of control and support subject choices Emphasise co-production of the initiative	Confirm and 'certificate' gains Retreat and diffuse information on 'success stories' for use at stage 1

chain is an iteration of program mechanisms. And mechanisms, recall, are the resources on offer within the program that, if triggered successfully, work their way into subjects' reasoning. From the vantage point of this article, the entire lower sequence should be understood as 'the program'. This representation, of course, is designed in vivid contrast to those research strategies that perceive and portray 'treatments' in glorious singularity.

With the basic model in place, it is then possible to extend it by postulating simple auxiliary theories about progress along the chain. We cash in its explanatory potential by making modest predictions about the need for integrity and continuity of the stages. For instance, a 'flows and blockages' theory would predict relative failure for those programs without comprehensive planning and implementation along the entire sequence. Meanwhile a 'loaded bases' theory would predict relative success for programs whose recruitment was secure so that they require little by way of publicity and promotion, so that most subjects have the head start of entering the scheme, so to speak, at stage 3. A 'joined-up thinking' theory would predict tensions in those programs having a marked division of labour between personnel responsible for promoting, recruiting, implementing and endorsing a scheme. A 'locals and cosmopolitans' theory would predict that individual practitioners rarely have equal control over, and skills in, dealing with the tacit and formal sections of the program pathway. A 'time is right' theory would predict that programs will struggle without an initial secular trend in climate of opinion in favour of its goals. A 'tipping point' theory would predict that program success requires a sufficient throughput of subjects at each stage for them to share responsibility for participatory progress.

More concrete, substantive examples of such hypotheses will be covered in the next two sections. The point here is simply to illustrate model dynamics, the potential explanatory power of envisioning programs as series of latent and manifest mechanisms of people processing.

Models simplify with a purpose, so let me now calm expectations about Figure 2 and the subsequent hypotheses. The most obvious point is that it is not to be applied mechanically. For instance, I am not claiming that eight phases, no more and no less, is the exact number of steps that must occur for programs to coalesce. Recruitment, selection and preparation may be automatic in some programs and exhaustive in other. Subject contact may be momentary or long-lived. Outcomes sought may be singular or multiple, they may be deep-seated or surface. Accordingly, the model may seem to elongate or compress the activities in any particular program.

But that, of course, is the point. The relative success of programs within the same family may often be explained because recruitment is simpler and more preconditions are met automatically (the loaded basis thesis). Youth mentoring provides an interesting example here. Enduring, prestigious programs such as the Big Brother/Big Sister

(BBBS) scheme in the US report high levels of outcome success compared with newly established programs—even those that seek to replicate the same core body of activities (DuBois et al. 2002; Grossman & Tierney 1998). A plausible explanation here is that 'flag raising', 'contrastive publicity' and the 'red carpet' are all firmly in place in a renowned intervention. The evidence here indicates that 'non-specifics' in BBBS are indeed lavishly and routinely primed, so much so that the program generates, wait for it again—a waiting list (Pawson 2006, p. 142).

The next point of clarification is to note that the model, whilst wide-ranging, is not claimed as a universal one. Its building blocks are drawn mostly from programs that provide new opportunities for individual clients, offered on a voluntary basis. These features mark its approximate domain. Middle-range theories are middle range in that whilst they attempt to 'conjoin different spheres of social behaviour', they do so in respect of 'delimited aspects of social phenomenon' (Merton 1967). Thus, Figure 2 would have to be reconfigured had we tried to make it apply, for instance, to community programs with collective actors, and also for mandatory (legislative) programs with generic subjects. And whilst I have suggested that some of its features are present in crime prevention programs, it would be wise to develop a different core model when the goal is to develop mechanisms for social control as opposed to precursors for opening up choice.

Finally on Figure 2, it might be worth clarifying again a crucial ontological point, namely and more plainly—'What's it all about?' How does it differ in content from the 'logic maps' and 'theories of change', which are commonplace in process-oriented evaluation? Again I stress the point about the process under consideration here being 'content-less' in that they try to map out preconditions for successful engagement to all manner of programs. So, unlike some logic models, the pathway specified is not about project management. It is not about applying for funding here, spending it here, or hiring staff here and recruiting subjects there. Neither is it about targets and auditing, about the way a set of intermediate outputs have to be met in order to lead to final outcomes. Nor is it about those sequential theories of behavioural change and the way interventions act on knowledge that percolates into attitudes and then, hopefully, into behavioural transformations. Rather, Figure 2 is about the routine practices of people processing. It is about the latent mechanisms that attract, recruit, hold and embed subjects into programs—in order that the manifest mechanisms may come into play.

Evaluating the power of invisible mechanisms

If the above model of program induction and conduction is correct, even approximately, then it has profound implications for the conduct of evaluation. Let us commence with a brief 'don't' before moving onto a medley of 'do's'.

The sweeping interlinkage of mechanisms described above *is* the program. Evaluation strategies that attempt to excise, minimise, partial out, or control for latent effects are missing the point. In social programs it is impossible to scrape away to the kernel agent for change, because change is always gradual and must be prompted gradually. The gallant attempts in clinical trials to eliminate human volition by double-blinding, the creation of waiting list controls and so on, should not be imitated in the evaluation of social programs. Such strategies merely camouflage the stakeholders' underlying choices, which are the genuine propellants of change. Hence, a waiting list from the point of view of this paper is not a null treatment, nor for that matter a performance target (for reduction). It is a choice, a moment when subjects ponder—should I stay or should I go? And the balance of such contemplation has profound effects for the progress of any intervention.

So much going on in the 'bow wave' and the 'side wash' and the wake of programs that it is not wise to iimagine that we can ever truly replicate their journeys. Mark Twain once said that history never repeats itself, but it rhymes. The same is true in the recurring echoes that are social programs. Under this model it is inevitable that programs are always implemented differently and this must be the starting point for a renewed evaluation agenda.

Implications

1 More research attention should be paid focusing on the stages in the above model—they and theories involved should become objects of inquiry in and of themselves. This would automatically bring to the surface the importance of invisible mechanisms. For instance, it would be quite possible to investigate the pros and cons of 'waiting lists' for a variety of different procedures, revealing no doubt different tipping points when their function changes from proving ground to detention bloc. Elsewhere in the model, the significance of 'quick wins' could be investigated across, say, regeneration programs, getting a measure of the importance of visible change for hard-to-reach populations. Such inquiry could be undertaken using primary research capturing these processes as they unfold, or by secondary review trying to piece together their imprint across a variety of programs and services.

2 More advantage should be taken of natural variation in program delivery. Such adaptation is obviously a feature of popular, widely instigated programs, which will bear the marks of the localities and times in which they are developed. It is also true of most 'corporately'-sponsored initiatives. So in health care systems the same innovation will be trialled across a number of

wards or units or hospitals. In regeneration programs a number of different 'partnerships' are often created to test-bed the latest ideas. In drug harm-prevention schemes there is often a roll-out across many schools and youth centres. As a consequence, we can always say that the routines of program development will always manufacture alternative ways of delivering the 'same' program. Such comparisons are more able to detect the significance of subtleties in program induction and throughput. How has access been managed in case 1 and 2 and did it involve that same subtle mix of the red-carpet treatment and rug pulling? Was there a difference in opportunities for program participants to learn from each other and influence the direction and content of the initiative? Such investigations could learn lessons from the methodology used in 'small n' comparative research and its notion of configurational causality (Ragin 1987).

3 More evaluation effort should be targeted at those interventions where latent forces loom largest and most controversially. Because their physiological, microbiological mechanisms of action are apparently inert, absent or unknown, much complementary and alternative medicine has been subject to charges of quackery. If, however, we take as the starting point that all interventions work by capturing hearts and minds as well as bodies, then a calmer approach to CAM evaluation can be contemplated (Bellavite et al. 2006).[6] Issues such as interpersonal, physical, non-verbal rapport and empathy (in whatever treatment) could be studied as change mechanisms in their own right. There would be ample room here for the golden rule of studying the 'same' program delivered in different ways. Homeopathy delivered from the high street (by, say, Holland & Barrett, UK's largest retailer of vitamin supplements and health food products) will have totally different dynamics than when developed in prolonged relationship with the registered practitioner.

4 More evaluation effort should be targeted on anomalies, outliers and unexpected consequences in explaining program progress. As we have seen from examples above, such investigations of program 'failure' have often been undertaken in the clinical field to great success. Lack of adherence to treatment was first studied in explaining variability across trials. This topic evolved with Donovan's studies (1995) discovering the phenomenon of patients' 'self-experimentation' to diminish side effects and in seeking the precise and optimal dosages that *they* deem effective. If we begin here, with the notion that subjects always undertake interventions in ways that 'seem right for them', another line of inquiry is opened. Inflexibility of provision may well be a general problem in what is predominately a top-down game. Subject inspired 'distortions' will always occur. The crucial task for process evaluation then becomes: how and to what extent can such forms of resistance be incorporated, in ways that allow the program to work to its original goals?

5 More longitudinal research should be conducted under the umbrella of evaluation. What happens upstream clearly conditions what occurs downstream. Most obviously, a poorly recruiting program or one that recruits the 'wrong' type of recruits is already on the high road to failure. But one suspects that flows and blockages occur throughout the life of a program, with equal significance for its fortunes. There are always refractory phases in the intervention pathway. Almost all practitioners in all fields will tell tales of gains lost when subjects perceive few signs of progress or when root problems return to enfold them. Program elements promoting reliance and stubbornness in pursuit of a goal might be as important as those designed to impart the skills and qualifications to achieve the said goal (Shiner et al. 2004). Our basic model tries to capture this process of continual reinforcement towards an objective and it could be used as a template to monitor the tempo of different groups in different conditions as they pick their way through a program.

These recommendations, of course, just begin to scratch the surface. The crucial point is that the coordination of a whole series of ideas and agents is required to create durable change. Evaluating program synchronicity will pay considerable dividends.

Notes

1 Consider for a moment a seemingly bizarre reading of the data in Figure 1. Since the waiting list control group registers considerable improvement over the norm, might this suggest that an efficacious and cost-effective way of reducing recidivism is to offer many such courses but not bother to get round to running them. Crazy? Well, yes in that such reasoning overlooks the self-selection of the volunteers. But, as we shall see later, 'decoy' interventions have become a real part of the policymakers' armoury.

2 I can assert this with some confidence, for it's high time I revealed that I was part of the research team.

3 These are known as 'active treatment comparisons'. Instead of randomly placing patients in treatment and control conditions, they are assigned to one of two treatments (e.g. cognitive versus behavioural). This ensures 'fair' comparison on a matched population.

4 He is describing a clinical experimental design, purportedly the ideal design, which operates with three groups, namely experimental, control and untreated. The first two receive a 'treatment' without knowing whether it is pill or placebo. The third is simply 'untreated'. The idea is that this strategy will perform the hat-trick—differentiating real effects from placebo effects from the null condition.

5 The first and most famous review of the placebo effect (Beecher 1955) made the claim that in the trials he examined there was a pattern whereby both experimental and control groups tended to show improvement. He estimated that one-third of the control group typically responds to placebos.

6 Bellavite and colleagues have produced a sequential, multiplicative and, thus what looks to my eyes, sound template for such evaluations: 'One can assume that in a homeopathic cure a complex interaction of these mechanisms occurs: (a) a small physical action of extremely low-dose remedy, (b) the activation of centres responding to "placebo effect" due to beliefs, expectations of the patient and (c) the endogenous healing mechanisms. If this is the case, the therapeutic effect is due not to the sum of these factors but their product and any procedure decreasing or shutting down one of them (as blinding undoubtedly does) may markedly affect homeopathic cure, much more than allopathic drug effect'.

References

Barclay, P, Buckly, P, Brantingdon, PJ, Brantingdon, PL & Whinn-Yeats, T 1997, 'Preventing auto thefts in commuter parking lots: a bike patrol in Vancouver', in R Clarke (ed.), *Situational crime prevention: successful case studies*, Harrow and Heston, Guilderland, New York.

Beecher H 1955,'The powerful placebo', *Journal of the American Medical Association*, vol. 159, no. 17, pp. 1602–1606.

Bellavite, P, Ortolani, R, Pontarollo, V, Benato, G & Conforti, A 2006, 'Immunology and homeopathy. 4. Clinical studies—part 2', *Evidence-based Complementary and Alternative Medicine (eCAM)*, vol 3, no. 4, pp. 397–409, <http://ecam.oxfordjournals.org/cgi/content/full/3/4/397>.

Chambliss, D 2002, 'Beware the dodo bird: the dangers of overgeneralization', *Clinical Psychology: Science and Practice*, vol. 9, no. 1, pp. 13–16.

Donovan, J 1995, 'Patient decision making. The missing ingredient in compliance research', *International Journal of Technology Assessment in Health Care*, vol. 11, no. 3, pp. 443–455.

Dracup, K & Meleis, A 1982, 'Compliance: an interactionist approach', *Nursing Research*, vol. 31, no. 1, pp. 31–36.

DuBois, D, Holloway, B, Valentine, J & Cooper, H 2002, 'Effectiveness of mentoring programs for youth; a meta-analytic review', *American Journal of Community Psychology*, vol. 30, no. 22, pp. 157–197.

Duguid, S 2000, *Can prisons work?*, University of Toronto Press, Toronto.

Grossman, J & Tierney, J, 1998, 'Does mentoring work? An impact study of the Big Brothers Big Sisters programme', *Evaluation Review*, vol. 22, no. 3, pp. 403–426.

Latey, P 1979, *Muscular manifesto*, Osteopathic Publishing, London.

Latey, P 2001, 'Placebo responses in bodywork', in D Peters (ed.), *Understanding the placebo effect in complementary medicine*, Churchill Livingstone, London.

Luborsky, L, Rosenthal, R, Diguer, L, Andrusyna, T, Berman, J, Levitt, J, Seligman, D & Krause, D 2002, 'The dodo bird verdict is alive and well: mostly', *Clinical Psychology: Science and Practice*, vol. 9, no. 1, pp. 2–12.

Luborsky, L, Singer, B & Luborsky, L 1975, 'Comparative studies of psychotherapies. Is it true that "everyone has won and all must have prizes"?', *Archives of General Psychiatry*, vol. 32, no. 8, pp. 153–155.

McGavock, H 1996, 'A review of the literature on drug adherence. Partnership in medicine taking', in *Taking medicine to best effect*, Royal Pharmaceutical Society of Great Britain, London, pp. 1–55.

Merton, R 1967, 'On sociological theories of the middle-range', Chapter 2 in R Merton, *On theoretical sociology: five essays old and new,* Free Press, New York.

Moerman, D 2002, *Meaning, medicine and the 'placebo effect'*, Cambridge University Press, Cambridge, England.

Parloff, M 1986, 'Frank's "common elements" in psychotherapy: non-specific factors and placebos', *American Journal of Orthopsychiatry*, vol. 56, no. 4, pp. 521–530.

Pawson, R 1996, 'Theorizing the interview', *British Journal of Sociology*, vol. 47, no. 3, pp. 296–314.

Pawson, R 2000, 'Middle-range realism', *Archive Européenes de Sociologie*, vol. XLI, pp. 283–325.

Pawson, R 2006, *Evidence-based policy: a realist perspective*, Sage, London.

Porporino, F & Robinson, D 1995, 'An evaluation of the reasoning and rehabilitation programme with Canadian federal prisoners', in R Ross & R Ross (eds), *Thinking straight*, Air Training Productions, Ottawa.

Ragin, C 1987, *The comparative method*, University of California Press, Berkeley, California.

Sallybanks, J 2001, *Assessing the police use of decoy vehicles*, Police research series paper 137, Home Office, London.

Shiner, M, Newburn, T, Young, T & Groban, S 2004, *Mentoring disaffected young people: an evaluation of 'Mentoring Plus',* Joseph Rowntree Foundation, York, England.

Smith, M, Clarke, R & Pease, K 2002, 'Anticipatory benefits in crime prevention', *Crime Prevention Studies Annual Review*, vol. 13, pp. 71–88.

Stewart, M 1995, 'Effective physician–patient communication and health outcomes: a review', *Canadian Medical Association Journal*, vol. 152, no. 9, pp. 1423–1433.

Strupp, H 1986, 'The nonspecific hypothesis of therapeutic effectiveness: a current assessment', *American Journal of Orthopsychiatry*, vol. 56, no. 4, pp. 515–20.

Tallman, K & Bohart, A 1999, 'The client as a common factor: clients as self-healers', in M Hubble et al., *The heart and soul of change: what happens in therapy*, American Psychological Association, Washington DC.

van Dulman, A & Bensig, J 2002, 'Health promoting effects of the physician patient encounter', *Psychology, Health & Medicine*, vol. 7, no. 3, pp. 289–300.

Vermeire, E, Hearnshaw, H & van Royen, P 2001, 'Patient adherence to treatment: three decades of research', *Journal of Clinical Pharmacy and Therapautics*, vol. 26, no. 5, pp. 331–342.

A keynote address to AES International Conference 2008, Perth

Improving project, program and policy performance in developing countries through managing for development results

This article seeks to answer five questions: (1) What is 'Managing for Development Results (MfDR)'? (2) What are the known conditions for the successful adoption of MfDR?, (3) Do these conditions apply to developing countries?, (4) Does MfDR produce better results in developing countries?, and (5) What are some of the emerging challenges in adopting MfDR? In noting that that MfDR is results-based management expressed in the language of development, the reasons for the adoption of the new term and the possible consequences of this are explored. While the perils of transferring management practices from one cultural context to another are noted, the Asian Development Bank's experience supports the view that the known conditions for successful adoption of MfDR generally apply to developing countries. Many developing countries also face special challenges not faced by rich countries and they need to deal with the particular cultural context that may support or impede successful adoption. Surprisingly little research has been conducted to test the relationship between MfDR and improved development results. However, the evidence presented in this article supports the view that applying the principles of MfDR can bring about significant improvements in project, program and policy performance. Paradoxically though, those countries that need the benefits the most are the least likely to be able to apply whole-of-government MfDR. Notwithstanding this, adoption where the conditions are favorable offers hope for better results from projects, programs and policies in the future.

R Keith Leonard

Scott Bayley

R Keith Leonard is Regional Director of the South Pacific Subregional Office at the Asian Development Bank, Suva. Email: <rkleonard@adb.org>

Scott Bayley is an Evaluation Specialist in the Operations Evaluation Department of the Asian Development Bank, Manila. Email: <sbayley@adb.org>

Introduction

Managing for Development Results (MfDR) is a firmly established part of the global development agenda.[1] The question for development agencies and their client countries is not whether they should adopt MfDR, but how. Heads of the multilateral development banks at the Second International Roundtable

on Managing for Development Results stated: 'We accord the highest importance to supporting countries in strengthening their capacity to better manage for development results'. In acknowledging a new partnership of shared responsibility they noted: 'Within this global partnership, countries must take the lead in managing their development and transition processes'. To assist in this process the heads of the multilateral development banks committed to 'strengthen partner countries' own institutions, systems, and capabilities to plan and implement projects and programs, report on results, and evaluate their development processes and outcomes, avoiding parallel donor-driven mechanisms'. An action plan was agreed (African Development Bank et al. 2004). This direction was strengthened further by the 2005 Paris Declaration on Aid Effectiveness.[2]

This article addresses the broad topic of improving project, program and policy performance in developing countries through MfDR by suggesting answers to five questions:

1 What is MfDR?

2 What are the known conditions for the successful adoption of MfDR?

3 Do these conditions apply to developing countries?

4 Does MfDR produce better results in developing countries?

5 What are emerging challenges in adopting MfDR in developing countries?

The evidence base for answering these questions is largely drawn from evaluation work carried out by the Operations Evaluation Department (OED) of the Asian Development Bank (ADB). The article ends by drawing conclusions.

What is 'managing for development results' (MfDR)?

MfDR is an approach to the delivery of aid (grants, technical assistance and concessional finance) that aims to produce better development results ('doing the right things and doing them right') through:

- being clear about the desired results in terms of impacts, outcomes and outputs
- understanding what must be done to achieve those results (spelling out the program logic or results chain)
- making management decisions on the basis of performance in producing results rather than the use of inputs.

The results being sought according to global consensus are 'sustainable improvements in country outcomes' (African Development Bank et al. 2004).

According to the ADB's *2004 Action Plan for Managing for Development Results*, the purpose of MfDR is to help managers answer three questions:

- Are we being effective?
- How do we know?
- How do we use this information to determine future action? (ADB 2004)

MfDR is results-based management (RBM) expressed in development terminology. Given this, the question arises, why invent a new term? Why not just call it RBM and get on with it? Clearly, any initiative to speed up the pace at which development takes place is highly desirable, and it is very positive that aid agencies, which are using taxpayer-provided money, and developing countries, which are spending it along with their own resources, are willing to be accountable for the results obtained. So, maybe we need not worry too much about what term is used. On the other hand, it can be asked whether the focus on the achievement of development results (in the developing countries) has the effect, whether intended or not, of deflecting attention from the application of the principles of RBM inside the aid organisation itself. If aid agencies are not practising RBM internally—that is, they do not 'practise what they preach' in their internal management—can they realistically expect to achieve better development results, and how credible is their message to their clients?

As suggested by this article, better development results can be achieved, to some extent at least, by following the principles of MfDR only in client countries. However, by not practising RBM internally, development agencies are transferring responsibility for achieving results principally to their clients and missing opportunities to improve their own effectiveness. Is this fair given the frequent difference in capacity? Probably not. Also, client feedback provided to OED as part of its evaluation work indicates that clients recognise that ADB does not always practise what it preaches and this affects their perception of ADB as a credible source of learning and policy advice.

Another concern regarding use of the term 'MfDR' is that it may restrict the focus to getting better results from aid flows rather than improving the results from a country's public expenditure in general. While improving the effectiveness of aid is important, for most developing countries, aid makes up a very small percentage of total public expenditure. The real gains are to be made in applying RBM to the entire government budget. If MfDR can act as a pilot for this to happen it will have achieved a great deal. However, there is a danger of isolation of RBM principles to the aid budget alone. We need to guard against this.

Turning now to the question at hand, MfDR/RBM can be broken down into three distinct activities—specifying expected results, measuring results and using that information to improve performance. MfDR only takes place when all three have been carried out.

1 **Specifying expected results.** Making explicit what results we expect to achieve before the policy, program or project is put in place usually

calls for results to be spelled out as activities to be undertaken, outputs to be produced, the outcomes that directly flow from having produced the outputs, and the broader impacts on people to which the project, program or policy will contribute. The main reason for breaking results down into this hierarchy is to account for differences in timing, measurability and attribution. At each level, measurable targets and indicators are set and baseline values determined. Additional steps are to: (a) spell out the data sources that will yield the required information for monitoring the achievement of results, and (b) the assumptions and risks that exist in the program logic of outputs leading to outcomes and impacts. What I have just described is the logical framework or design and monitoring framework as we term it in the ADB (ADB 2007b).

2 **Performance measurement**. The targets and indicators are tracked as a means of determining progress towards achieving the expected results. It is also desirable to monitor whether the program logic assumptions are continuing to hold true and risks, whether identified ex-ante or not, have occurred thus requiring mitigation. Collectively, these activities comprise monitoring. Evaluation goes beyond monitoring to determine why results were or were not achieved, and to identify good practice and lessons for the future.

3 **Performance management**. This involves the use of the data provided by performance measurement. 'Managing for results involves using information to improve decision-making and steer country-led development processes toward clearly defined goals' (African Development Bank et al. 2004).

What are the known conditions for successful adoption of MfDR?

What does the literature tell us about the necessary conditions for MfDR's adoption? OED recently completed an evaluation of ADB's progress in adopting MfDR (ADB 2007c) and this contains a literature review that identified seven success factors that need to be present in an organisation for MfDR to be successfully adopted. These are:

- strong support from the senior leadership
- staff with the necessary skills, receiving appropriate training
- an organisational culture that focuses on achieving results
- supportive staff management practices and incentives
- sufficient operational capacity and resources
- appropriate management of change process for introducing MfDR
- suitable business systems and processes in place.

Lessons emerging from international experience (taken from the same OED study) are summarised in Figure 1 under three categories: (1) promoting favourable implementation conditions, (2) developing a performance measurement system, and (3) using performance measurement information.

FIGURE 1: LESSONS FROM ADOPTION OF RESULTS-BASED MANAGEMENT

Promoting favorable implementation conditions

- Customise the results-based management system to users' needs rather than adopting someone else's approach.
- Take time and maintain momentum—be patient and persistent as it can take years for success.
- Connect performance management to the business plan and budget to give meaning.
- Align management systems to support implementation, but use existing systems and information whenever possible.
- Provide adequate financial and human resources.
- Take care over the location within the organisation of the stewardship of the performance management process—do not isolate it.
- Pilot projects can be a useful approach.
- Developing a performance management culture is essential but changing an established culture is not easy; thus it takes time.
- Changes are required in the notions and locations of accountability in public sector organisations.
- Senior-level leadership is necessary for successful implementation and this should be a shared responsibility at both political and management levels.
- Broad-based participation fosters support for implementation.
- Training and education build success by contributing to culture change and helping to overcome a lack of experience and expertise.
- Communicate the purpose of RBM throughout the organisation and to external stakeholders.

Developing a performance measurement system

- Keep it simple and use a manageable number of indicators, but use a balanced set of measures to ensure an adequate performance assessment.
- Consider three performance measurement system essentials: design, use and updates.
- Understand the three types of performance measures: key results indicators, performance indicators and key performance indicators.
- Clearly define key terms and concepts.
- Use logic charts to help define expected results, indicators and risks.
- Align performance measures with accountability and decision-making authority—individuals should be held accountable for what they can influence and they must understand the system and how they can influence results. Ultimately, there should be shared accountability for results.
- Generate credible performance information (valid and reliable) and subject this to some form of independent checking.
- Use frameworks to achieve balance in measurement systems.
- Performance standards and targets are essential for measurement and accountability—benchmarking can be a useful approach while the use of baseline data is also highly recommended.

Using performance information

- Demonstrate the use of performance information—it must be used and seen to be used by others. Visible use by top managers is particularly important.
- Effective management requires performance information, capacities to use that information, and incentives to act on the performance information—without an incentive to act, performance information and capacities are useless.
- Evaluation complements performance measurement by helping explain why targets were not met or were exceeded and giving guidance on how to improve programs.
- Incentives can be used to foster support—the most successful RBM systems are non-punitive.
- Performance reporting is needed for decision-making—a vital flow of information is required. However, the information reported must be credible and presented in a way that facilitates its use.
- Learn, review and adjust performance measurement systems.

Do the conditions for successful adoption of MfDR apply to developing countries?

The perils of transferring management practices from developed to developing countries are well known (see for example ADB 2008 and Schick 1998). Since MfDR is the product of a 'Western' management tradition, we need to be cautious about saying that these lessons apply, or that they apply to the same extent, in the developing country context. Perhaps there are other factors that are more important in terms of success? Is MfDR even relevant to many developing countries? Schick (1998) provides a sobering assessment in respect of the relevance of the New Zealand reform model to developing countries.

> I believe there are important preconditions for successfully implementing the new public management approach and these should not be ignored by countries striving to correct decades of mismanagement. In contrast to those who take the position that managerial deficiencies should be the driving factor in determining the suitability of these types of reforms, I argue that they should be deterring factors. The greater shortcomings in a country's established management practices, the less suitable the reforms.

To help address the question of what factors are important in the developing country context, we can use the experience of OED's efforts since the early 1990s to build evaluation capacity in seven of ADB's developing member countries. This draws on an OED evaluation of that experience up to 2001 (ADB 2001) supplemented by further experience in the People's Republic of China. The main lessons from this experience are:

- It proved more complex and time-consuming to build a functioning results-based monitoring and evaluation system than first thought.
- High-level commitment to performance evaluation is essential—this should be clearly demonstrated by allocation of sufficient budget and staff, the status of the activity within the

government and individual institution, and the existence of at least one influential 'champion' for results management.

- There are a number of preconditions for success in building evaluation capacity: (a) a commitment within the government to accountability, (b) management decisions are not made on political grounds, (c) an incentive structure exists that encourages public sector employees to operate in the public interest, and (d) stability in staffing.
- Building capacity in results monitoring and evaluation requires incentives and training at all levels: field, project management unit, province and central agency levels.
- Establishing a results-based monitoring and evaluation system is a means to an end, not an end in itself. It has a cost. Benefits are only obtained when the results are used in key budget, investment and policy/strategy decision-making processes. The design of results monitoring systems should not assume that information produced by these will be used. Rather, specific design features must be included to encourage and formalise use. Opportunities need to be captured for providing results-based monitoring and evaluation information to policymakers.
- Capacity building is a long-term process.

A World Bank Independent Evaluation Group study produced similar findings to that of ADB (Mackay 2007). It notes: 'First and foremost is that *substantive demand from the government is a prerequisite to successful implementation*' (emphasis in original). The study identifies three dimensions of success as: (1) utilisation of monitoring and evaluation information, (2) good-quality monitoring and evaluation information, and (3) sustainability. In citing Chile as an example of what 'success looks like', Mackay (2006) says that 'in particular it is the intensive utilization of the monitoring information and evaluation findings which the M&E system produces … [that] are used by the Finance ministry for its resource allocation decisions with the budget process, and to impose management and efficiency improvements on sector ministries in the programs for which they are responsible'.

Clearly, there are a number of similarities among the lessons emerging from the international literature and OED's experience in seven of ADB's client countries. The importance of leadership, the time required for success, the need to create demand for results-based monitoring and evaluation information, and the need to align with existing budget and other systems stand out. However, there are clues to possible important differences. One is the availability of resources to support MfDR. These are likely to be much scarcer than in developed countries. Another difference is the extent to which the necessary preconditions for success identified in point (3) above—namely, a demonstrated commitment to public sector accountability, limited political interference in management decisions, an incentive structure that encourages public servants to operate in the public interest, and a degree of stability in staffing. Most of these can safely be assumed to be true in developed countries, at least most of the time. The same cannot be said for many developing countries. Another significant difference between developing and developed countries is the level of capacity, particularly at mid and lower levels of organisations.

Another factor that should be considered is that of culture. The literature on RBM emphasises the importance of establishing a culture change that is supportive of managing on the basis of results. Even the most superficial observation reveals that developing countries cover a wide range of cultures and cultural traditions and that these differ greatly from that prevalent in the developed countries where RBM originated. Does national culture matter when introducing MfDR and can we measure the likely effect? In reviewing eight studies that sought to cluster countries on the basis of cultural variables, Ronen and Shenkar (1985) concluded that this was a valid process that reflected real differences. In his seminal work, Hofstede (1980, 1991) ranked countries on the basis of four value dimensions (later five): (1) power distance, (2) individualism/collectivism, (3) masculinity/femininity, (4) uncertainty avoidance, and (5) short-term/long-term orientation. Cultures also vary in terms of the way they establish truth and reality; for example, the relative importance placed on facts, figures and logic; as opposed to feelings, intuition and spirituality (Schneider & Barsoux, 1997). Without entering into details of the many studies on national culture and the implications of this for management, the reality of differing cultural values among nations certainly supports a conclusion that the introduction of MfDR in developing countries will have to confront challenges that may differ significantly from the countries in which the approach originated. While the globalisation of ideas and education may have narrowed the cultural differences among nations somewhat since much of this research was done, enough differences remain to demand caution when introducing MfDR in developing countries.

Another difference between developed and developing countries is discussed by Caiden and Wildavsky (1990) when seeking to explain budgetary practices in rich versus poor countries—in particular, the need for frequent budget changes in the latter. They argue that the occurrence of the problem in many poor countries of very different races and culture suggests that common problems (rather than differences) are at play. In their view 'the basic cause of the phenomenon in low-income countries is extreme and extensive uncertainty which, when combined with severe scarcity of financial resources, narrows the time horizons of top officials to two or three months or less'. This is hardly a context in which MfDR will

flourish. Paradoxically, therefore, it appears that the countries that need MfDR the most may be the least likely to be able to adopt it. Conversely, the conditions may be more suitable in countries where greater certainty and stability prevails.

Does MfDR produce better results in developing countries?

Whether the adoption of MfDR produces better results is a critical question, but perhaps surprisingly, there is a lack of empirical research addressing it. On the other hand, perhaps this is not so surprising as it is generally considered axiomatic that MfDR is a 'good thing'. As evaluators we support the principles of MfDR wholeheartedly but we should be rigorous in assessing the performance of MfDR in practice. Clearly, MfDR has a cost and the benefits should exceed the costs to make the effort worthwhile. In this section four pieces of evidence are presented that support the view that MfDR can contribute to better results: (1) the improvement in the performance of ADB's policy-based lending operations, (2) the overall trend towards improved performance of ADB lending operations and the convergence in the performance of ADB's ordinary capital resource and concessional lending operations, (3) shared characteristics of successful projects in ADB, and (4) an example of anecdotal evidence from a client. Certainly, the evidence is not definitive, but it is suggestive that positive gains for development are possible from applying the principles embodied in MfDR. However, the question of whether the value of the benefits exceeds the costs remains to be answered.

The case of ADB's experience with policy-based lending

As well as funding investment projects, ADB provides budget support loans to encourage policy reform—these operations are known as program or policy-based loans. The loans are generally disbursed in one to three tranches following compliance by the client with a number of agreed reform conditions. Since introducing program lending in 1978 until 31 December 2006, ADB approved 184 such loans to 31 client countries for a total amount of $24 billion (ADB 2007d). The success of the 101 completed operations with program completion or independent evaluation assessments has varied widely over the years but with noticeable trends as shown in Figure 1. Overall, 51per cent were rated 'successful', 46 per cent 'partly successful' and 3 per cent 'unsuccessful'.

The 2007 ADB evaluation study (ADB 2007d) sets out to identify the factors that favour or hinder success. While not a specific finding of the study, it is reasonable to suppose that the steady decline in the performance of program loans in the 1980s to a nadir of 0 per cent successful in 1991 focused the attention of managers on the need for corrective action and certainly, the success rate has climbed dramatically since then, being maintained around the 70 per cent mark. The evaluation highlighted a number of factors associated with a lack of success—among others these included:

- over-ambition within the proposed time frame; overly complex designs
- deficiencies in the program logic
- a lack of clarity on expected results and a failure to communicate these
- the importance of political economy dimensions
- absence of, or deficiencies in, measurement indicators for expected results (particularly quantitative indicators)
- a lack of reliable and timely information on which to base reform program management
- a failure to monitor results following program termination
- a lack of dialogue around the emerging evidence on outcomes
- the need to link policy reform to budgets via medium-term expenditure frameworks.

Taking these success factors into account has contributed to better performance of ADB's policy-based lending.

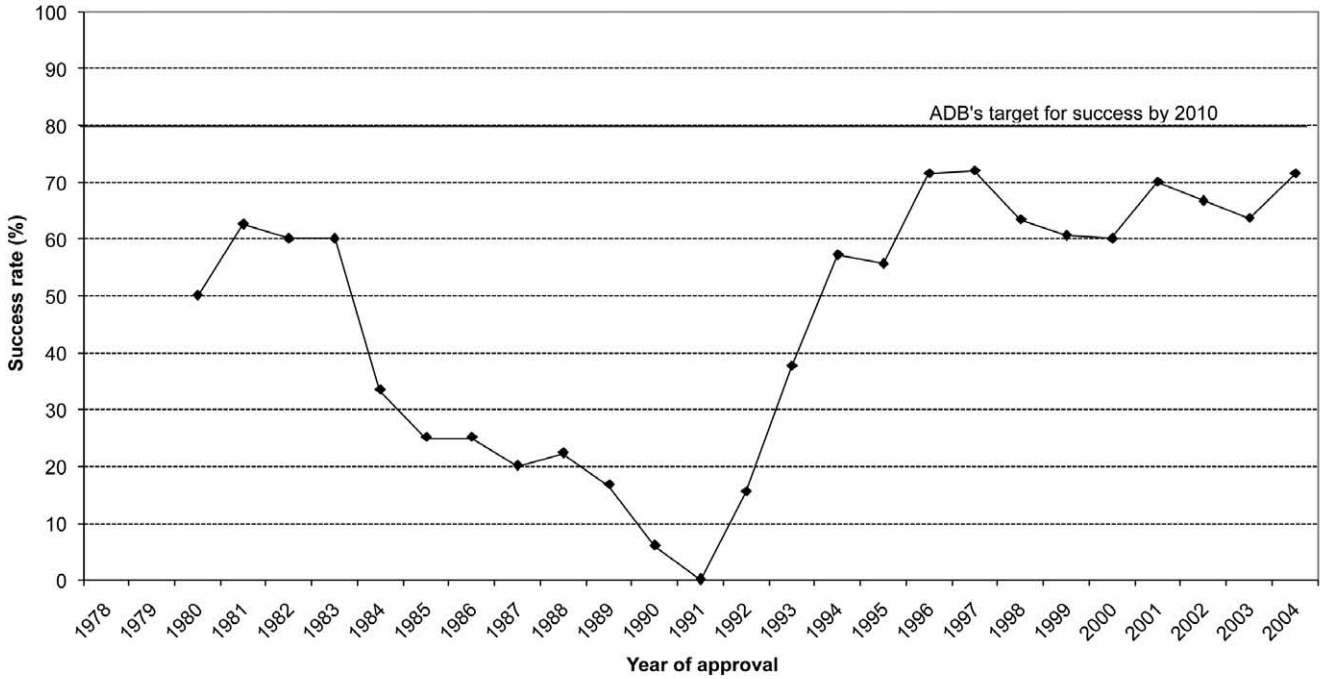
The case of trends in the overall performance of ADB operations

The second piece of evidence in support of the contention that application of the principles underlying MfDR can contribute to better development results is provided by the trend to improved overall project performance (not just policy-based lending) and the convergence of the success rate of ADB's lending under its Ordinary Capital Resource and Asian Development Fund windows (the latter provides concessional funding to the poorer countries). The success rates are shown in Figure 2.[3] The trend is for improved performance and a convergence of performance by funding source.

OED's 2005 Annual Evaluation Review (ADB 2005) confirmed the trend to improved performance after controlling for the variables of country and sector (projects are more likely to be successful in some countries and sectors than others). The study found a clear association between design quality and project success. Based on an assessment of 2001 projects with post-evaluation reports, it found that 85 per cent of projects with minimal design issues were rated successful, compared with 52 per cent with moderate design issues, and 17 per cent for those with major design issues. Further, it found that the proportion of projects with minimal design issues rose from 18 per cent for those approved in the period 1976–1979, to 42 per cent for 1980–1989, and to 65 per cent for 1990–1995. The report concludes that 'the design of ADB projects has progressively improved, an indication of continued

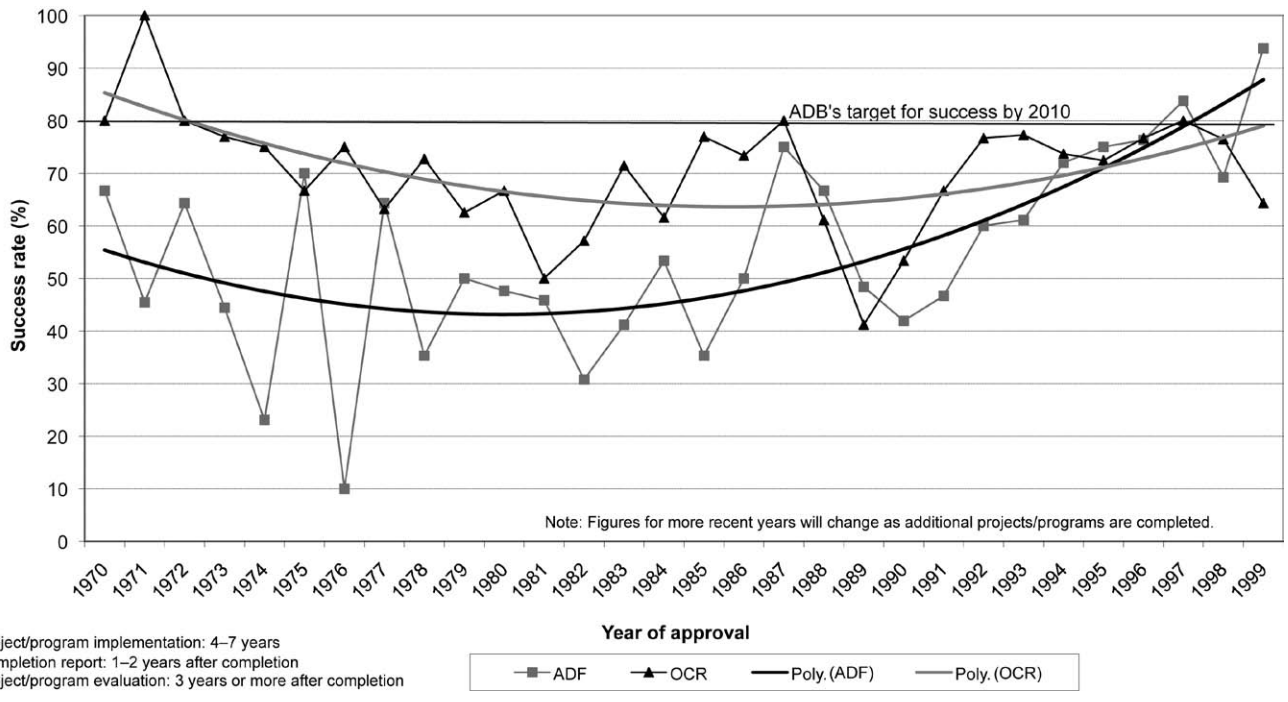
FIGURE 2: TRENDS IN PROGRAM RATINGS BY YEAR OF APPROVAL

Percentage of program loans rated successful based on a three-year moving average



Source: Operations Evaluation Department's database

FIGURE 3: PROPORTION OF SUCCESSFUL OPERATIONS BY SOURCE OF FUNDS



Project/program implementation: 4–7 years
Completion report: 1–2 years after completion
Project/program evaluation: 3 years or more after completion

Source: Operations Evaluation Department's database

learning and the effectiveness of initiatives aimed at strengthening project quality' (ADB 2005). Having said this, as noted later, the designers of ADB-funded projects are still having difficulty in defining expected results at the impact and outcome levels, and a significant part of the judgement of success may have been on the basis of output rather than outcome or impact achievement.

A recent OED study on the two most recent replenishments of the Asian Development Fund (ADB 2007e) calls the degree of convergence achieved between the performance of Ordinary Capital Resource and Asian Development Fund operations 'remarkable', given the generally lower levels of capacity in the less-developed countries that receive concessional funding. It opines a number of factors may have contributed to this trend including: improved capacity in the poorer countries following repeat operations, the gradual delegation of responsibility for project administration to ADB's resident missions, and the smaller size of Asian Development Fund operations (they are on average three times smaller), which may allow them to receive relatively more attention. In addition to the factors identified by the evaluation study, one can note the introduction of strategic planning by ADB in the early 1990s and the adoption of a set of strategic development objectives that brought greater clarity to the outcomes being sought from ADB's operations: use of the logical framework became prevalent and then mandatory during the decade, and an action plan produced by a task force on project quality was implemented to improve project performance. These initiatives embodied many of the principles of MfDR.

Characteristics of successful ADB projects

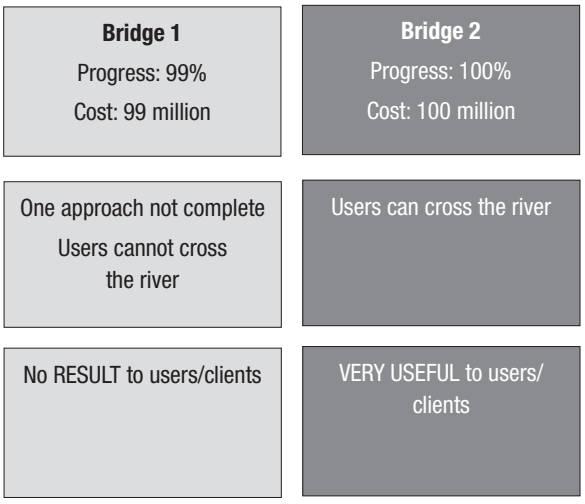
A third strand of evidence in support of the contention that adoption of MfDR produces better results is provided by looking at the characteristics of successful projects. OED's 2006 Annual Evaluation Review (ADB 2006b) sought to identify common factors contributing to successful projects in five core sectors (road, power, water supply/sanitation, education, and irrigation/drainage sectors). Among the 10 common factors, the following were noted:

- A striking feature of successful projects was their ability to learn lessons from past experience and incorporate these lessons into the design of subsequent projects.
- Executing agencies and ADB staff were able to identify and solve problems during the implementation of successful projects.
- Flexibility by ADB in allowing appropriate design changes, a proactive stance in solving problems, and making approvals in a timely manner were characteristics of successful projects.
- The presence of baseline information appeared to be associated with more successful outcomes in the road sector.
- Over-optimism in the specification of expected results were often associated with less successful projects in the power and road sector operations.

The experience of the Bangladesh Local Government Engineering Department

The fourth element of evidence is the experience of the Local Government Engineering Department in Bangladesh, as presented in a workshop funded under an ADB regional technical assistance (ADB 2006a).[4] The presentation compared the assessment of success based on output delivery versus outcome achievement. The example was of two bridges with associated approach roads. In one situation despite 99 per cent of the outputs being delivered, one part of an approach road was not complete. No benefits flowed so the outcome of 'increased and improved accessibility and mobility of people and of agricultural/non-agricultural goods' was not achieved. The very different assessment of performance when outcome attainment is taken into account is illustrated in Figure 4. This simple example belies a revolution in thinking that underpins it. Very different decisions will be made when success is judged on outcomes rather than outputs, and these decisions can make a dramatic difference to people's lives and hence development.

FIGURE 4: ASSESSING OUTCOME VERSUS OUTPUT ACHIEVEMENT



Source: A presentation by M Zulyaminayn, Monitoring Engineer, Local Government Engineering Department, Bangladesh as presented in a seminar held in Bangkok, 13–14 November 2007 under the auspices of Asian Development Bank Regional Technical Assistance 6306: *Mainstreaming managing for development results in support of poverty reduction in South Asia.*

What are emerging challenges in adopting MfDR?

Accepting that the application MfDR can produce better development results, this section deals with a number of challenges we have encountered in ADB in seeking to apply it. It is suggested that many of these are relevant to developing countries in this part of the world as (1) the experience derives from designing and implementing projects in the Asia-Pacific region, and (2) a substantial majority of ADB's staff are from the region.

Specifying and agreeing expected results is not as easy it seems

A starting point for successful MfDR is to be able to specify expected results at various levels (outputs, outcomes and impacts), along with targets, indicators, baseline levels, data sources for monitoring and evaluation, and assumptions and risks. As previously mentioned, all ADB projects are required to have a design and monitoring framework where the required information is spelled out. Detailed guidelines exist and these have been modified several times over the years (ADB 2007c). ADB has invested considerable resources in providing training to staff to ensure the high quality of design and monitoring frameworks. Internal quality control systems have been put in place. Notwithstanding this, the quality remains disappointing. According to assessments carried out for OED, frameworks for projects approved in 2006 were only marginally better than those approved in 2000.

OED's experience in assessing the quality of design and monitoring frameworks has revealed two things: (1) it's hard to get the 'experts' to agree on what constitutes good quality, and (2) it's very hard to get project, program and policy proponents to come up with a good-quality frameworks—statements of expected results are at the wrong level or unclear (particularly at the outcome and impact levels), the program logic is not logical, indicators are missing for some results areas or are not relevant, targets and baselines are not given, data sources are poorly specified, and statements of assumptions and risks are often inadequate (including the inclusion of so-called 'killer risks or assumptions' that have a high probability and would doom the project, program or policy to failure if they occur (risks) or do not hold true (assumptions).

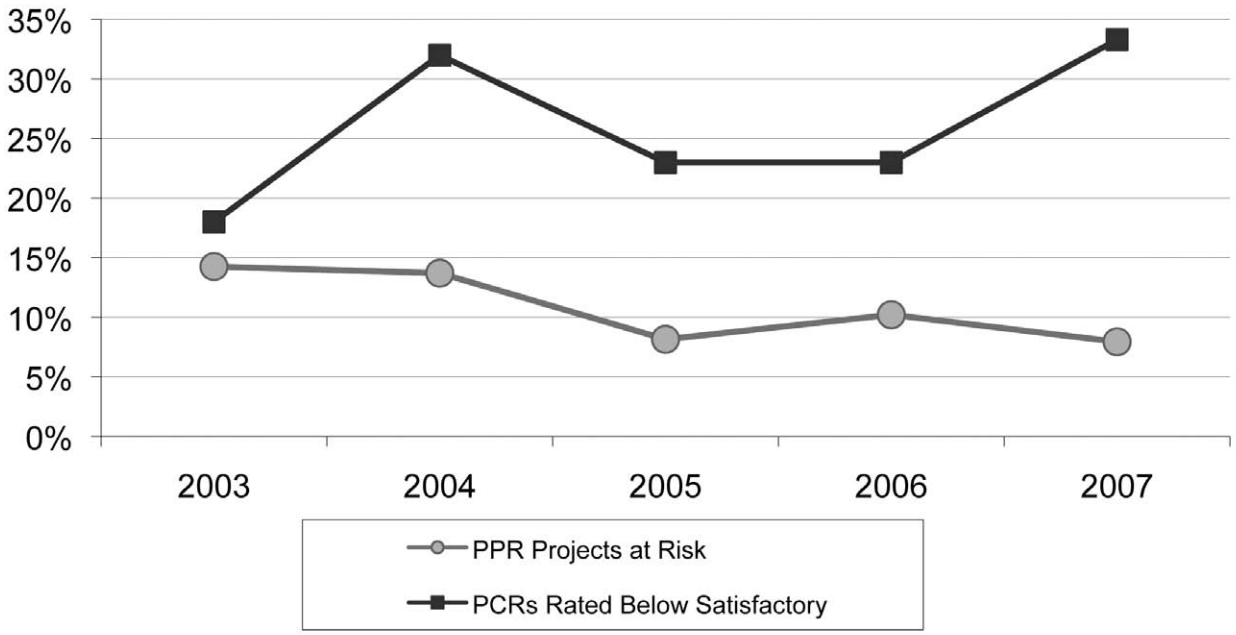
Our experience shows the following may contribute to the difficulty in getting good-quality design and monitoring frameworks:

- Despite instructions to the contrary, design and monitoring frameworks tend to be produced as an adjunct to the formulation process of the new project, program or policy rather than being used as (1) a design tool for conceptualising and analysis, (2) the basis for engaging with project stakeholders during formulation, and (3) an effective means of communicating thinking on results and the program logic to key decision-makers **throughout** the formulation process. Previously, design and monitoring frameworks were mostly prepared as an afterthought following completion of design simply because it was mandatory to include one in the documentation. Given this reality, those responsible for project or policy formulation often see the design and monitoring framework as a necessary evil rather than something useful.
- Projects, programs and policies are too frequently submitted for funding with a very unclear definition and understanding of the problem to be solved (or in the worst cases, none at all). Not surprisingly, it is impossible to specify expected results if there is a lack of clarity about what the problem is, and what its underlying causes are. As the Cheshire Cat said to Alice in Wonderland in a well-known children's story, 'If you don't know where you are going, any road will take you there'. ADB requires project design teams to construct a problem tree but this appears to be done infrequently.
- Even where the problem is reasonably well defined, there is generally a lack of quantified baseline information. Often the need for baseline information is acknowledged by saying this will be collected as part of implementation, but frequently this does not happen. Setting achievable targets is pretty well impossible in the absence of baseline information.
- If they can get away with it, people would rather not be held accountable for specific achievements so there is a natural reluctance to establish the basis by which one might be held accountable in the future.
- There is frequently a conflict between a stated commitment to the achievement of development results and the political reality of multiple agendas and objectives (both within the development agency and the government). A focus on the achievement of development results calls for simple designs with a limited number of outcomes, whereas the political reality of multiple agendas often results in projects, programs and policies with multiple expected outcomes and a consequent complexity and, in ADB's experience, reduced chances of success.

Problems in performance measurement during implementation

Beyond the problems with ex-ante specification of expected results, there are also problems in measuring performance during project implementation—in ADB the tool for this is known as the project performance report (PPR). This is designed to be a tool that alerts managers to underperforming projects that need remedial attention. It does this by rating projects as 'problem projects' (or not) and 'potential problem projects' (those not currently a problem but showing early warning signs that they are heading in that direction) with the combination

FIGURE 5: LESS THAN SUCCESSFUL/SATISFACTORY RATINGS IN PROJECT PERFORMANCE AND COMPLETION REPORTS



Source: Operations Evaluation Department's database

of the two categories being known as 'at-risk projects'. Every completed project has an evaluation of performance in the form of a project completion report (PCR). These rate projects as highly successful, successful, partly successful or unsuccessful. If the project performance monitoring system is working as intended, there should be a close relationship between the proportion of projects rated 'at risk' in the final PPR before closure and the PCR. As shown in Figure 5, this is far from the case. This phenomenon is known as the 'deathbed' conversion whereby apparently satisfactory projects become partly successful on completion.

OED's Annual Report on 2006 Portfolio Performance examined this problem in some detail (ADB 2007a). Using a random sample of 65 projects under implementation, it found that 44 per cent lacked the basis to measure progress towards the achievement of impact and outcome while 11 per cent even lacked the basis for determining implementation progress. Whereas according to the 65 PPRs, 98.4 per cent of projects were rated satisfactory or better, OED's reassessment was 25.4 per cent—the corresponding figures for partly satisfactory were 1.6 per cent for PPRs and 28.6 per cent for OED's assessment. Interviews were conducted with project staff and 50 per cent of project officers doubted the system's ability to track progress towards the achievement of development objectives. The following factors were identified as contributing to this reality:

- poor-quality design and monitoring frameworks rendering monitoring difficult, if not impossible
- deficiencies in the use of the PPR, including a lack of sector expertise among project staff, excessive optimism, and insufficient candour in the reporting and ratings
- systemic inconsistencies in the PPR system
- user unfriendliness of the system
- contradictory incentives as a result of using the ratings generated by a system designed as an early warning tool as an overall measure of success of the portfolio. For the purpose for which the system was designed, it is desirable to have around 25 per cent of the portfolio identified as being 'at risk' to know where remedial attention is required. If used as a measure of success, as few as possible should be rated 'at risk'.

Organisational challenges in adopting MfDR

In this section, selected results are presented from a recent OED evaluation of MfDR in ADB (ADB 2007c). This evaluation included a staff perceptions survey. When asked to rank the relative importance in ADB of seven factors shown in the international literature to be critical for successful adoption of MfDR, 74 per cent of 772 staff rated senior leadership support as 'very important'. The next two highest ranked criteria were human resources practices, and staff skills and training. Results are shown for some of the questions on senior leadership support and human resources practices and incentives, along with some questions looking at whether staff believe MfDR is here to stay and is seen as very important by ADB. The results should give pause for thought to those promoting MfDR at the institutional level.

Getting senior management to lead and be accountable[5]

Table 1 shows that only 33 per cent of respondent staff agree that ADB's management actively supports the results agenda in a tangible way (e.g. by demanding information on results). Of particular note, only 16 per cent of international staff below director level and somewhat less than a third at director level or above agreed with the statement. National officers and administrative staff were somewhat more positive although they had a high level of 'don't know' answers. Clearly, ADB management is going to have to do a lot more to convince staff that it is serious about MfDR and not just paying lip-service to it.

Table 2 shows the results of staff views on whether ADB management is held accountable for the achievement of development results. Here, administrative staff were the most positive with 67 per cent agreeing with the statement while international staff were much more negative in their views. More effort is required to demonstrate that achieving development results is taken seriously by the organisation, particularly to international staff.

Getting the organisational incentives right

As shown in Table 3, ADB is a long way away from having the right human resource systems in place to

TABLE 1: STATEMENT: ADB'S MANAGEMENT ACTIVELY SUPPORTS THE RESULTS AGENDA IN A TANGIBLE WAY

	Disagreed	Neutral	Agreed	Don't know	Total N
	% of responses				
Overall result	21	29	33	17	811
International staff director level and above	36	33	31	0	42
International staff below director level	43	30	16	11	227
National officers	16	26	40	18	152
Administrative staff	7	30	43	20	328

TABLE 2: STATEMENT: ADB MANAGEMENT IS HELD ACCOUNTABLE FOR ACHIEVING DEVELOPMENT RESULTS

	Disagreed	Neutral	Agreed	Don't know	Total N
	% of responses				
Overall result	28	18	45	9	882
International staff director level and above	57	24	19	0	42
International staff below director level	62	15	15	7	227
National officers	22	24	45	9	152
Administrative staff	8	16	67	9	328

encourage a results focus. At one extreme, only 11 per cent of international staff below director level agreed with the statement.

Staff were even more negative about the proposition that current incentives in the organisation encourage staff to manage for development results—only 19 per cent of director level and above international staff thought so while below the director level, almost none of the 227 respondents were of that opinion. See Table 4.

Generating belief

Table 5 indicates that a significant minority of international staff believes MfDR is just another

TABLE 3: STATEMENT: HUMAN RESOURCE SYSTEMS MOTIVATE STAFF TO FOCUS ON RESULTS IN THEIR WORK

	Disagreed	Neutral	Agreed	Don't know	Total N
	% of responses				
Overall result	46	22	27	4	811
International staff director level and above	64	14	21	0	42
International staff below director level	69	17	11	4	227
National officers	47	22	26	5	152
Administrative staff	29	29	38	4	328

TABLE 4: STATEMENT: THE CURRENT INCENTIVES ENCOURAGE STAFF TO MANAGE FOR DEVELOPMENT RESULTS

	Disagreed	Neutral	Agreed	Don't know	Total N
	% of responses				
Overall result	54	25	14	7	801
International staff director level and above	64	17	19	0	42
International staff below director level	78	16	2	k4	227
National officers	51	25	15	9	152
Administrative staff	37	33	21	9	328

TABLE 5: STATEMENT: MANAGING FOR DEVELOPMENT RESULTS IS A FAD

	Disagreed	Neutral	Agreed	Don't know	Total N
	% of responses				
Overall result	33	31	26	11	866
International staff director level and above	36	21	43	0	42
International staff below director level	30	27	37	5	227
National officers	34	31	26	9	152
Administrative staff	33	37	16	15	328

TABLE 6: STATEMENT: ADB ACHIEVING DEVELOPMENT RESULTS IS MORE IMPORTANT THAN ACHIEVING DISBURSEMENT AND LENDING TARGETS

	Disagreed	Neutral	Agreed	Don't know	Total N
	% of responses				
Overall result	38	23	29	11	904
International staff director level and above	55	21	24	0	42
International staff below director level	66	14	16	4	227
National officers	42	25	24	9	152
Administrative staff	16	30	38	16	328

development fad while around a fifth were neutral. Only around a third disagreed with the statement. More staff disagreed (substantially more in the case of international staff at any level and national officers) that achieving development results was more important than achieving disbursement targets in ADB than those that agreed with the proposition (see Table 6).

While the survey also revealed some positives, the overall message is that ADB has a long way to go to meet the conditions generally shown to be necessary for the successful adoption of MfDR. The study also concluded that progress in adopting MfDR was broadly similar among multilateral development banks. The results confirm that the adoption of MfDR can require a long-term commitment.

Conclusions

The evidence presented supports the view that application of the principles of MfDR can improve the performance of projects, programs and policies in developing countries and it can do so to a very significant extent. However, MfDR is not a panacea for past poor performance. Its adoption will not address the underlying causes of poor governance, corruption, and the ineffective and inefficient management of inputs (including staff along with physical and financial resources). Paradoxically, those countries that are the least developed and which most require the benefits MfDR are less likely to have the necessary preconditions for success.

This seems like a fairly dismal conclusion but all is not lost. While whole-of-government adoption of MfDR may be beyond the less-developed countries, a start can and should be made on putting it in place where conditions are favorable within individual departments or institutions and at subnational levels of government. The existence of a 'champion for change' with power and influence is an essential ingredient for success at whatever level.

Even where the preconditions are more favourable for successful adoption of MfDR, expectations about the pace of change need to be realistic and based on a clear understanding of the particular context—one size is most unlikely to fit all, so customised solutions are needed. This is where initiatives such as the ADB-supported by country-led MfDR Community of Practice[6] can play a very effective role in helping practitioners customise the approach to the particular circumstances they are facing and to learn from the experience of others.

For their part, development agencies such as the ADB need to do significantly more to apply the principles of MfDR to their internal workings.

Notes

1 As outlined at <www.mfdr.org>, MfDR had its origins in the establishment of the Millennium Development Goals and the International Conference on Financing for Development in Monterrey, Mexico (2002), and was concretized in the International Roundtable on Measuring, Monitoring, and Managing for Results (2002), and Second International Roundtable on Managing for Development Results, in Marrakech, Morocco (2004).

2 For further information on the *2005 Paris Declaration on Aid Effectiveness*, refer to <http://www.aidharmonization.org/ah-overview/secondary-pages/editable?key=205>.

3 Figure 2 shows project success rate by year of approval. Most projects approved after 2000 are still ongoing. The results shown for later years rely on fewer projects and more on self-evaluations by operational departments than independent evaluation. The completion of less well-performing projects is frequently delayed so these may not yet show up in the results. Hence, the success rates for later years may be lower as more delayed projects enter the database.

4 The website for the technical assistance is available at <http://www.adb.org/Documents/Events/2007/Mainstreaming-MfDR/default.asp>.

5 In the tables that follow international staff are professionals recruited in the international marketplace from among ADB's member countries; national officers are professionally qualified staff recruited and working in their home country—a majority is in resident missions; administrative staff are also working in their home country and are predominantly Filipinos in ADB's headquarters in Manila.

6 More information is available at <http://cop-mfdr.adb.org>.

References

ADB 2001, *Selected technical assistance for strengthening evaluation capacity in developing member countries*, Asian Development Bank, Manila.

ADB 2004, *2004 Action Plan for Managing for Development Results*, Asian Development Bank, Manila, <http://www.adb.org/MfDR/actionplan/default.asp>.

ADB 2005, *The 2005 Annual Evaluation Review*, Asian Development Bank, Manila.

ADB, 2006a, *Mainstreaming MfDR in support of poverty reduction in South Asia,* Asian Development Bank, Manila, <http://www.adb.org/projects/project.asp?id=39601>.

ADB 2006b, *The 2006 annual evaluation review*, Asian Development Bank, Manila.

ADB 2007a, *Annual report on 2006 portfolio performance,* Asian Development Bank, Manila.

ADB 2007b, *Guidelines for preparing a design and monitoring framework,* Asian Development Bank, Manila, <http://www.adb.org/Documents/Guidelines/guidelines-preparing-dmf/default.asp>.

ADB, 2007c, *Managing for development results in the Asian Development Bank: A Preliminary Assessment*, Asian Development Bank Manila.

ADB 2007d, *Policy-based lending: emerging practices in supporting reforms in developing member countries*, Asian Development Bank, Manila.

ADB 2007e, *Special evaluation study on the Asian Development Fund viii and ix operations*, Asian Development Bank, Manila.

ADB 2008, *Performance evaluation report, Mongolia: Governance Reform Program*, Manila.

African Development Bank, Asian Development Bank, European Bank for Reconstruction and Development, Inter-American Development Bank, and World Bank 2004, *Joint Marrakech Memorandum* and *Annex 1: promoting a harmonized approach to managing for development results* and *Annex 2: action plan on managing for development results*, <http://www.mfdr.org/2ndRoundtable.html>.

Caiden, N & A Wildavsky 1990, *Planning and budgeting in poor countries*, Transaction Publishers, New Brunswick, New Jersey.

Hofstede, G 1980, *Culture's consequences: international differences in work-related values*, Sage, Beverley Hills, California.

Hofstede, G 1991, *Cultures and organizations: software of the mind*, McGraw Hill, London.

Mackay, K 2006, *Institutionalization of monitoring and evaluation systems to improve public sector management*, World Bank Independent Evaluation Group, Washington DC.

Mackay, K 2007, *How to build M&E systems to support better government*, World Bank Independent Evaluation Group, Washington DC.

Ronen, S & Shenkar, O 1985, 'Clustering countries on attitudinal dimensions: a review and synthesis', *Academy of Management Review,* vol. 10, no. 3, pp. 435–454.

Schick, A 1998, 'Why most developing countries should not try New Zealand reforms', *The World Bank Research Observer,* vol. 13, no. 1, pp. 123–131.

Schneider, SC & JL Barsoux 1997, *Managing across cultures*, Prentice Hall, Hemel Hempstead, England.

REFEREED ARTICLE
Evaluation Journal of Australasia, Vol. 8, No. 2, 2008, pp. 28–34

Participatory research in challenging circumstances

Lessons with a rural Aboriginal program

The evaluation of a crime prevention program for Aboriginal boys in a rural setting (Tirkandi Inaburra Cultural and Development Centre) faced a number of challenges, such as geographic distance, a modest budget, and the need for culturally appropriate and sustainable methods that capture a multitude of program effects while minimising the burden of data collection.

Literature suggests the value of participatory methods. However, would this approach work in these circumstances? Would it threaten the probity of the evaluation?

This article describes a case study of participatory methodology used in challenging circumstances and considers the factors that contributed to its successful implementation. These included mutual respect, willingness to work together, high capacity within Tirkandi Inaburra, funding-body support for participatory approach, a common concern for the program's mission, and strategies for ensuring probity.

It is recommended that funding bodies accommodate the resource implications involved in developing relationships and changes to research plans that can be necessary for genuine participation.

Catherine Spooner

Saul Flaxman

Colleen Murray

Catherine Spooner (top left) is a Senior Research Fellow at the National Centre in HIV Social Research, University of New South Wales, Sydney.
Email: <c.spooner@unsw.edu.au>

Saul Flaxman (top right) is a Research Associate at the Social Policy Research Centre, University of New South Wales, Sydney.
Email: <s.flaxman@unsw.edu.au>

Colleen Murray (bottom) is Executive Officer of the Tirkandi Inaburra Cultural and Development Centre, Coleambally, New South Wales. Email: <cmurray@tirkandi.org.au>

Introduction

There is substantial literature on the value of participatory research methods, but few demonstrated case studies from which researchers can draw lessons. This article describes the application of a participatory approach with an Aboriginal organisation in challenging circumstances.

The Tirkandi Inaburra Cultural and Development Centre (TI) is an Aboriginal community-controlled program that aims to reduce the risk of program participants becoming involved in the criminal justice system. Program participants are 12- to 15-year-old Aboriginal boys from across a catchment area that encompasses central southern New South Wales (NSW). TI is funded by the NSW Attorney General's Department (AGD) as a demonstration project and opened its doors to the first intake of

participants in January 2006. In 2006, the AGD sought tenders from organisations: (1) to assist TI in its conceptualisation and development of data collection tools for ongoing monitoring; and (2) to conduct an outcomes study that would assess the effectiveness of TI in achieving its objectives for participants, particularly in relation to education, social integration and involvement in the criminal justice system. The research project is important for both program development and for funding decisions. Information from the evaluation aims to contribute to decisions on funding of the program and the possible expansion of the program to other sites in NSW.

In developing a research protocol to meet the AGD's needs, a number of challenges were identified. Foremost of these challenges were two issues related to geographic distance. First, the Chief Researcher was based in Sydney, 600 kilometres from the site of the TI program. This suggested that on-site data collection and liaison with program staff would be a costly endeavour. Second, the homes of the program participants were geographically dispersed across communities up to 600 kilometres from the TI site, so follow-up would be difficult.

The evaluation faced numerous additional challenges. One was ensuring the evaluation was culturally appropriate. The program participants and most staff were Aboriginal, whereas the Chief Investigator was not. The program participants were young (12 to 15 years), geographically dispersed and generally had a history of problem behaviours, so engagement, obtaining informed consent and follow-up was likely to be difficult. The research brief stipulated that evaluation and monitoring should continue after the evaluation was completed, so the methods and instruments needed to be sustainable. The evaluation needed to capture a multitude of possible program effects while minimising the burden of the data collection on program participants and TI staff members. Finally, the research budget was small, so monetary solutions to these challenges (e.g. travel and an on-site research team) were not possible.

It was envisioned that a participatory research approach in which the program being evaluated was active in all stages of the research would help to address the challenges listed above. Involving TI staff in research design would hopefully ensure that the evaluation was appropriate and captured the domains of interest. In addition, data collection by TI staff would reduce costs and embed the data collection methods in the program's routine.

There is an extensive literature on the benefits of participatory research methods (outlined below), particularly with marginalised communities (Power 2002) and with Aboriginal communities (Fisher & Ball 2003) suggesting that it was an appropriate approach. Nevertheless, in planning the research, it was not certain that participatory methods would be feasible or effective. A number of problems could arise. For example, would the geographic distance between the researchers and the program site impede authentic participation? Would program staff be willing and able to participate in the research? Would involving program staff in data collection with minimal on-site supervision affect the probity of the evaluation?

In this paper, we report on how a participatory research approach was implemented in a situation that was not highly conducive to such a collaborative approach. We identify strategies we used to make the approach work as well as factors that fortuitously aided the process. While there is substantial literature on the value of participatory research and possible problems, there are few reports on the use of participatory research in challenging conditions. We go on to consider how researchers and funding bodies can reduce the risk associated with participatory research in such conditions.

Participatory research

The following section contains a brief overview of participatory research literature.

What is participatory research?

The term 'participatory research' (PR) can refer to a wide range of research practices that might also be described as collaborative, cooperative or empowering. As discussed by Patton (2002), the fundamental element of PR is a 'commitment to involving people in the setting being studied as co-inquirers' (Patton 2002). While actual methods can vary, a number of principles of genuine PR have been identified by Patton. These include:

- involving participants in learning inquiry logic and skills
- participants owning the research—participation is real, not token
- participants work as a group, the researcher supports group cohesion and functioning
- all aspects of the research are conducted in ways that are meaningful and understood by the participants
- the researcher acts as a facilitator, collaborator and resource; participants are co-equal
- the researcher recognises and values the views and skills of the participants and works to help participants recognise their own and each other's expertise
- status and power differences between researcher and participants are minimised.

These principles guide both what is carried out during PR and the manner in which it is performed.

Reasons for using participatory approaches

PR is a break away from the tradition of 'experts' conducting research *on* people (Levin 1993). Levin identified three benefits of PR:

- to increase the use of the findings by those involved
- to ground the data in participants' perspectives
- to mobilise social action.

Other researchers have identified additional benefits:

- to teach inquiry logic and skills (Patton 2002)
- cost-effectiveness (Chambers & Mayoux 2004).

Issues in conducting participatory research

While PR has potential benefits, it also has potential problems. For example, Lennie (2006) identified the following issues:

- ensuring stakeholder representativeness
- managing conflicting agenda
- naive assumptions about participation leading to empowerment
- the need for time and resources to develop relationships.

Other issues that require consideration include:

- ensuring genuine participation
- probity
- intellectual property.

These are discussed briefly below.

Ensuring genuine participation

An often-discussed issue in relation to PR is the level of participation. A number of models have been published to represent the range of possible levels of participation from non-participation to tokenism to empowerment (Cornwall & Jewkes 1995). In the 1960s Sherry Arnstein conceptualised eight levels of participation from non-participation to tokenism to citizen power (Arnstein 1969). More recently, and drawing on Biggs (1989), Cornwall and Jewkes (1995) identified four levels of ascending participation; contractual, consultative, collaborative and collegiate. The former types of participation they describe as being 'shallow', while the latter types are identified as 'deeper' participation in which the researchers relinquish more control to the participants (Cornwall & Jewkes 1995). The aim of genuine participation then is to move towards deeper participation.

Chambers and Mayoux (2004) argued that it cannot be assumed that participants will benefit or be empowered by participatory approaches. They claimed that in some instances participants will actually bear a cost or burden as a result of their participation. For example, people's enthusiasm might inadvertently cause them to reveal sensitive information that could be harmful to themselves or their community (Chambers & Mayoux 2004). Further, they noted that power relations need to be considered but not just between the researchers and the researched, but also between participants themselves (Chambers & Mayoux 2004). This can occur when one participant dominates the process, intimidating other participants and thereby skewing representativeness of the research. Similarly, Fisher and Ball (2003) identified the possibility of a backlash if a community perceives the 'participatory' aspect to be superficial.

Ensuring stakeholder representativeness

As discussed by Farrington (1997), participation is not only about *depth* (how people are involved) but also about *breadth* (who is involved). Some advantages of broadly involving stakeholders are that the research may be perceived as being more legitimate, it may build social capital and can more readily overcome conflicts. However, the disadvantages may include a proportionally larger expenditure of time and resources as well as the potential to generate more conflict if the representation is too broad (Koontz & Johnson 2004). De Lancer Julnes (2001) emphasises that diverse stakeholder representation is essential to successful participatory evaluation but cites a number of potential problems that can arise when consultation is wide and stakeholders are broadly involved. De Lancer Julnes warns of the potential for certain voices to dominate others, and for narrow agendas to be promoted to the detriment of others.

Managing conflicting agendas

Gregory points out that while power almost always underpins participatory evaluation, it is often ignored (Gregory 2000, cited in Lennie 2006). The conflicting agenda and perspectives that individual stakeholders and groups bring to an evaluation can impede its effectiveness. The more people are involved in a project, and the more deeply they are involved, the more ideas and agenda there are to be managed. Accordingly, communication is even more important in larger research projects with large numbers of stakeholders, such as the TI evaluation.

The need for time and resources to develop relationships

Lennie (2006) argued that there is a need for sufficient time and resources to be invested for the participatory method to succeed. This is particularly important in the early stages of the process. Some participants might not be familiar with the method or might think that the evaluation could have ramifications for funding or their employment (Lennie 2006). Further, Letiecq and Bailey (2004) explained that when the non-Indigenous researchers are physically separated, these investments are crucial for successful PR.

Probity

Some literature has thrown doubt on the validity of participatory approaches, claiming that it is biased and political (Cornwall & Jewkes 1995). However, others have argued that the participatory approach is more rigorous because it acknowledges more overtly its intentions to all participants—researchers and observers (David 2002). Either argument could

have substance, depending on how the approach is implemented.

Intellectual property

The participatory methodology raises the question of who owns the intellectual property from the research. Chambers has commented that, 'in good PRA [participatory rural appraisal] practice there is a tradition that the data—the maps, matrices and diagrams—belong to those who created them'. (Chambers & Mayoux 2004, p. 16). The principle that intellectual property generated from the study of a particular group should remain under the control of that group is linked to the recognition of the tradition of power relations, and the need to address this (Fitzgerald 2005). Guidelines for research with Indigenous peoples affirm this principle (for example, see the Australian Institute for Aboriginal and Torres Strait Islander Studies' (AIATSIS) *Guidelines for Ethical Research in Indigenous Studies* (2000). This can be challenging if there is a tension between the Indigenous participants' requirement for empowerment, and the obligations of the researcher to report results.

PR with Indigenous peoples

PR has been used extensively with Indigenous populations around the world, including Australia, New Zealand, the United States of America and Canada (Caldwell et al. 2005; Couzos et al. 2005; Fitzgerald 2005; Kowal, Anderson, & Bailie 2005; Letiecq & Bailey 2004). PR that is consistent with the principles outlined above is recommended by Indigenous research guidelines such as those of the AIATSIS (2000), the Aboriginal Health and Medical Research Council (AHMRC 2007) and the National Health and Medical Research Council (NHMRC 2003). These guidelines also emphasise the need to acknowledge historical disempowerment by ensuring there be a benefit to the community, reciprocity between the researcher and researched, and most importantly, community control over data produced.

Current literature notes the importance of identifying the tradition of power differentials between the researcher and the researched and identifies the need to lessen the impact of this. For example, Fitzgerald argued that 'the goal is not to replicate these power relationships but to challenge and change them and work towards *a 'cross-cultural competency'*" (Fitzgerald 2005, p. 18). Participatory approaches can address an Indigenous community's need for empowerment through research. Conversely, the non-Indigenous academic relies on the Indigenous participant(s) to allow 'cross-cultural competency' to be realised.

As noted by Chambers (1997), PR with remote or rural communities is difficult given that most academics are urban-based. The physical distance between the researchers and the researched can create problems, even if they are recognised and attempts are made to compensate by adopting a participatory methodology. A project that found logistical constraints was conducted on an American Indian reservation, which was located approximately 300 miles from the university base of the academic researchers (Letiecq & Bailey 2004, p. 354). While visits were made to the research site, they were infrequent due to constraints on time and resources and additionally restricted by adverse weather conditions. Letiecq and Bailey (2004) noted that while contact via telephone and email allowed communication with local project staff, these technologies could not substitute for face-to-face interaction. Such interaction, they argued, is essential for developing relationships 'above and beyond professional relationships' (Letiecq & Bailey 2004, p. 354) or to meet the requirement of 'reciprocity' that such cross-cultural projects demand.

Thus, the literature suggests that PR can be a valuable approach, but is susceptible to a range of possible problems. Given the challenges to the research project discussed here, the feasibility, value and probity of the PR approach were uncertain. In this article, we describe how we used a participatory approach despite geographic distance and limited resources, the benefits obtained by using a participatory approach, and the factors that contributed to the feasibility, value and probity of the approach.

Participatory methods in the current study

In this section, the processes, outputs and problems of the participatory approach in our study are outlined. To reiterate, the obstacles to the approach working included the small budget for travel for face-to-face meetings and on-site observation, and the geographic distance between the research team and the program being evaluated. Program participants were a marginalised group: young, rural, Aboriginal boys with behavioural and literacy problems. The program was run by an Aboriginal community with whom the research team had no prior relationships. The factors that contributed to the success of the approach in this situation are discussed below.

Processes and outputs

The current study included a participatory approach from its inception. An outline for a study was presented to the funding body with a participatory approach proposed for developing the research. Consequently, while the study was imposed on TI by its funding body, the funding body allowed a participatory approach to be used in the study.

In terms of breadth of participation, consultation occurred with the TI Board, TI Manager, TI staff, the funding body (AGD) and TI participants ('boys'). Consultation on research design with the parents of the boys could not be conducted due to resource constraints. Consultation with the boys by research staff was minimal for the same reason, and was restricted to discussion with a group of boys about a draft questionnaire. Otherwise, the research team was reliant on the TI Manager and staff to

inform the project regarding the parents' and the boys' perspectives.

Consultations with others were conducted via workshops and meetings. In these meetings, the key domains of interest to the evaluation, the research methods, instrument design and the study results were discussed. All participants contributed to discussion, and decision-making was based on group consensus. Meetings necessitated research staff travel by aeroplane to a regional airport and a drive of one-and-a-half hours from Griffith to a purpose-built facility in a rural setting.

Further (and frequent) collaboration occurred between the researchers and the TI Manager and Case Manager via telephone and email.

Specific immediate outcomes of the collaborations were:

- identification of important domains for measurement
- instrument design
- TI staff involvement in obtaining informed consent
- TI staff involvement in data collection
- TI staff priming participants for follow-up
- interpretation of the study results
- a joint presentation by the Chief Investigator and the TI Executive Officer at the Australian Social Policy Conference
- this journal article.

Anticipated longer term outcomes are:

- TI will be able to monitor and evaluate itself better after the evaluation is completed
- TI will own the report when it is released and be more receptive to the report as a tool for improvement than if they had had no involvement in the study design and implementation
- TI will understand the implications of the report better than if they had no involvement in the study design, analysis and reporting.

Problems

While the participatory approach has many benefits, as identified in the literature and above in the current study, there were two outcomes that were problematic, both of which relate to the problem of needing sufficient flexibility and resources to incorporate genuinely the ideas of the participants.

First, there was a tendency for the evaluation to expand beyond the resources because the more people involved, the longer the wish list. While it is the role of project management to ensure the research plan does not exceed the budget, this was a difficult task because the budget was small relative to the research requirements. Consequently, the research team undertook a substantial amount of work that was not funded.

These changes did not happen just at the beginning of the project. There was a tendency to make changes to the research plan after ethics approval had been obtained. This meant repeated requests to the ethics committee for approval for the changes, which took a large amount of (unfunded) time. Again, it is the role of the project manager to ensure a project adheres to its plan. However, to have done so rigidly would have been to ignore the feedback from TI participants about the need for changes, which would have damaged relationships as well as the study quality.

While the problem of distance was addressed via frequent telephone and email contacts, it cannot be said that distance had no impact. In comparison with previous experience with a similar evaluation in which the Chief Investigator had a desk on-site (Spooner, Mattick & Noffs 2001), there was much less opportunity in this study to develop relationships with participants and staff and to observe program implementation. Further, monitoring of program staff in the collection of consent and data was hampered. An on-site evaluator could have easily checked that procedures were being implemented correctly. As off-site evaluators, we were reliant on telephone and email to monitor progress. However, program staff were busy with program implementation and often unavailable to talk with the research team. As noted above, distance and limited budget also prevented extensive involvement of the program participants and other stakeholders (in particular parents/guardians and people who referred boys to the program, e.g. school teachers and juvenile justice officers) in study design.

A potential problem related to probity. For budgetary reasons, it was impossible for the research team to recruit participants (explain the study and obtain consent) or to collect the baseline and post-test data. These tasks were performed by the TI Case Manager and not supervised by the researchers. While data collection by TI staff was consistent with notions of ensuring sustainability of the data collection process, it does raise concerns about the probity of the study.

However, for a number of reasons, we argue that probity has not been sacrificed. First, the declared interest of the TI Manager, and the impression of the researchers, was that TI had a stake in ensuring probity as they wanted to know the truth about the effectiveness of their program and wanted to ensure that the evaluation was taken seriously by the funding body and others. Further, the evaluation instruments involved multiple-choice questions that allowed the participant himself to circle the answer that best described his feelings and/or situation. TI staff did not conduct all data collection. An independent observer was employed to observe program implementation for four days and follow-up interviews were conducted by independent researchers. Any discrepancy between data collected by TI and data collected by researchers would have suggested a potential problem with probity. Finally,

a representative of the funding body, who has a clear interest in probity, was one of the participants in the research process (attending meetings on research design). So, it is unlikely that the participatory approach has compromised the probity of this study.

Factors contributing to the feasibility of this participatory project

Given the challenges to participatory research outlined above, it was quite possible that the approach might not have been a productive one. Poor relationships could have developed with one side blaming the other for problems. In addition, miscommunications could have resulted in major mistakes, conflicting agenda might not have been resolved and program staff might have been unable to contribute. Yet, the approach worked well, as indicated by the successful achievement of tasks (research design and data collection). The factors that appear to have contributed to this success were:

- mutual respect
- willingness to work together
- support from the funding body for the participatory approach
- high capacity to collaborate within the organisation—the Manager and Case Manager already had well-developed skills relevant to designing, implementing and analysing research
- common concern for the plight of Indigenous youth and the necessity to review a model objectively, which might lead to breaking the cycle of involvement for Indigenous youth in the criminal justice system
- creation of a research consortium with one local person (who understood rural issues) and one Indigenous researcher (who understood cultural issues).

While the Chief Investigator was able to create a research consortium that facilitated a participatory approach, the researchers had no prior knowledge that the other facilitating factors existed. In other words, the approach was a risky one.

Conclusions

The success of the participatory process was not evaluated formally. Rather, this article represents the shared learnings from the process, from the combined perspectives of the researchers and the TI Manager. While the research literature suggested that the participatory approach is particularly difficult when resources and geographic distance are barriers to building relationships, our experience has been that these barriers need not be insurmountable. With an organisation willing and able to be active research partners, and researchers willing to go the extra mile, PR can be successful. However, while we managed in these difficult conditions, and many others have done so before us, this was the result of dedicated individuals rather than adequate funding.

In the current environment of government tendering for research projects in a competitive manner, researchers are under pressure to submit proposals that aim to achieve ambitious objectives with minimal budgets. The time frames for preparing tenders is generally short (weeks rather than months) and the research team often has no prior knowledge of, or relationship with, the organisation to be 'researched'. Accordingly, the researcher often does not know the capacity or willingness of the organisation to participate. Our recommendations for researchers tendering for projects are to:

- conduct background research on the organisation or group to be researched in order to identify their willingness and capacity to collaborate
- ensure the budget enables travel and time for building relationships and working together
- ensure the budget and time frame have flexibility to incorporate changes, as suggested by research participants.

It is recognised that time and budgetary pressures make the implementation of these recommendations difficult. To support researchers in these tasks, we recommend that funders of research:

- recognise the value of participatory research
- allow funding and timelines that allow for genuine collaboration and participation
- participate—be part of the team.

Acknowledgements

We would like to thank the NSW Attorney General's Department for funding this project, Liz West from the funding body for being so flexible and supportive of participatory research, the TI Management Board, staff and participants of TI for allowing us to work with them, and Elise Trask for assistance in the production of this article.

References

Aboriginal Health and Medical Research Council of New South Wales (AHMRC) 2007, *Ethics and research*, viewed 21 January 2008, <http://www.ahmrc.org.au/Ethics%20and%20Research.htm>.

Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS) 2000, *Guidelines for ethical research in Indigenous studies*, viewed 21 January 2008, <http://www.aiatsis.gov.au/__data/assets/pdf_file/2290/ethics_guidelines.pdf>.

Arnstein, SR 1969, 'A ladder of citizen participation', *JAIP,* vol. 35, no. 4, pp. 216–224.

Biggs, S 1989, 'Resource-poor farmer participation in research: a synthesis of experiences from nine national agricultural research systems. OFCOR comparative study paper 3', paper presented at the International Service for National Agricultural Research, The Hague, The Netherlands.

Caldwell, JY, Davis, JD, Du Bois, B, Echo-Hawk, H, Erickson, JS, Goins, T et al. 2005, 'Culturally competent research with American Indians and Alaska

natives: findings and recommendations of the first symposium of the work group on American Indian research and program evaluation methodology', *American Indian and Alaska Native Mental Health Research: The Journal of the National Center,* vol. 12, no. 1, pp. 1–21.

Chambers, R 1997, *Whose reality counts?, Putting the first last*, Intermediate Technology Publications, London.

Chambers, R & Mayoux, L 2004, 'Reversing the paradigm: quantification and participatory methods', paper presented at the EDIAIS Conference on 'New directions in impact assessment for development: methods and practice', University of Manchester, England.

Cornwall, A, & Jewkes, R 1995, 'What is participatory research?', *Social Science & Medicine,* vol. 41, no. 12, pp. 1667–1676.

Couzos, S., Lea, T., Murray, R., & Culbong, M 2005, 'We are not just participants—we are in charge: The NACCHO ear trial and the process for Aboriginal community-controlled health research', *Ethnicity and Health,* vol. 10, no. 2, pp. 91–111.

David, M 2002, 'Problems of participation: the limits of action research', *International Journal of Social Research Methodology,* vol. 5, no. 1, pp. 11–17.

de Lancer Julnes, P 2001, 'Does participation increase perceptions of usefulness? An evaluation of a participatory approach to the development of performance measures', *Public Performance and Management Review,* vol. 24, no. 4, pp. 403–418.

Farrington, J 1997, 'Farmers' participation in agricultural research and extension: lessons from the last decade', *Biotechnology and Development Monitor,* no. 30, p. 1215.

Fisher, PA & Ball, TJ 2003, 'Tribal participatory research: mechanisms of a collaborative model', *American Journal of Community Psychology,* vol. 32, nos 3–4, pp. 207–216.

Fitzgerald, T 2005, 'Cross-cultural research principles & partnerships: experiences from New Zealand and Australia', *Management in Education,* vol. 19, no. 1, pp. 17–20.

Koontz, TM & Johnson, EM 2004, 'One size does not fit all: matching breadth of stakeholder participation to watershed group accomplishments' (author abstract), *Policy Sciences,* vol. 37, no. 2, pp. 185–204.

Kowal, E, Anderson, I & Bailie, R 2005, 'Moving beyond good intentions: Indigenous participation in Aboriginal and Torres Strait Islander health research', *Australian and New Zealand Journal of Public Health*, vol. 29, no. 5, pp. 468–470.

Lennie, J 2006, 'Increasing the rigour and trustworthiness of participatory evaluations: learnings from the field', *Evaluation Journal of Australasia,* vol. 6, no. 1, pp. 27–35.

Letiecq, BL & Bailey, SJ 2004, 'Evaluating from the outside: conducting cross-cultural evaluation research on an American Indian reservation', *Evaluation Review*, vol. 28, no. 4, pp. 342–357.

Levin, B 1993, 'Collaborative research in and with organizations', *Qualitative Studies in Education,* vol. 6, no. 4, pp. 331–340.

National Health and Medical Research Council (NHMRC) 2003, *Values and ethics: guidelines for ethical conduct in Aboriginal and Torres Strait Islander health research*, viewed 21 January 2008, <www.nhmrc.gov.au/health_ethics/human/conduct/guidelines/_files/e52.pdf>.

Patton, MQ 2002, *Qualitative research and evaluation methods*, 3rd edn, Sage, London.

Power, R 2002, 'Participatory research amongst marginal groups: drug users, homeless people and gay men', *Drugs: Education Prevention and Policy*, vol. 9, no. 2, pp. 125–131.

Spooner, C, Mattick, RP & Noffs, W 2001, 'Outcomes of a comprehensive treatment program for adolescents with a substance-use disorder', *Journal of Substance Abuse Treatment,* vol. 20, no. 3, pp. 205–213.

REFEREED ARTICLE
Evaluation Journal of Australasia, Vol. 8, No. 2, 2008, pp. 35–41

The political context of evaluation: what does this mean for independence and objectivity?

Anne Markiewicz

Anne Markiewicz is the Director of Anne Markiewicz and Associates. Anne has been operating her consultancy for the past 12 years, specialising in developing monitoring and evaluation systems and conducting evaluation projects. Email: <anne@anneconsulting.com.au>

Evaluation takes place within a political and stakeholder context that has benefits for the relevance and usefulness of the evaluation. However, the politicised context also presents challenges in preserving evaluator independence and objectivity, with potential adverse consequences for the credibility of the evaluation. This article proposes that evaluators need to recognise and negotiate these challenges effectively to ensure that a quality evaluation results.

Introduction

Evaluation takes place within a political context. It is also increasingly influenced by the variety of positions presented by actively involved stakeholders. The political and stakeholder influenced context of evaluation is a reality that brings with it many benefits for the relevance and the use of evaluation findings. However, this context also presents a number of challenges for the evaluator. In particular, there are challenges for the evaluator in preserving their independence and objectivity in a politicised context, with potential adverse consequences for the credibility of the evaluation.

Many evaluators have experienced undue influence from commissioners of evaluations or key stakeholders involved in the process. This influence can affect each stage of the evaluation process but is particularly highlighted during the formulation of evaluation findings or recommendations. It is proposed that evaluators need to negotiate such challenges effectively, potentially armed with evaluation standards to support their case. A credible evaluation is necessary to enable both commissioners of evaluations to make informed decisions and evaluators to produce quality work. The knowledge, skills and experience of the evaluator in attaining balance between the impact of political and stakeholder influences and interests and delivering credible evaluation findings is critical to the future of evaluation and its utility.

The challenges presented by the political and stakeholder context of evaluation do raise the longstanding paradigm wars between scientific realists and social constructionists. The former group of evaluators tend to uphold concepts of objectivity and independence in evaluation, while the latter group of evaluators view themselves as negotiators of different social realities (Taylor & Balloch 2005, p. 1). Here it is proposed that it should be the common endeavour for all evaluators along the continuum of scientific realism and social constructionism to produce credible evaluations that reflect some degree of objectivity and independence from political and stakeholder influence.

This article commences by examining briefly the political and policy-driven context within which evaluation takes place. It neither attempts to provide a

theoretical analysis of the policy context, nor debate whether the political context should impact on evaluation. It is considered to be a given that the political and policy context does have considerable impact on evaluations. Rather, this discussion provides a summary of the challenges and implications for the evaluator of working within a political context. This is followed by an examination of the stakeholder context in evaluation. Furthermore, the adoption of utilisation-focused and participatory models that involve stakeholders in the evaluation process are supported. It is suggested that, as with the political context, there has to be some reconciliation between the influence of the participating key stakeholders in the evaluation and the credibility of the evaluation process and its findings.

Concepts of credibility and trustworthiness in evaluation are briefly discussed next, and this is followed by consideration of the concepts of independence and objectivity in evaluation. Finally, consideration is given to the role of the Australasian Evaluation Society in developing practice standards to support the work of evaluators in their complex negotiations with commissioners and stakeholders in regard to issues such as independence and objectivity.

The challenge for evaluation highlighted in this article is thus:

The process of adopting a politically grounded, policy-relevant and participatory approach to evaluation whilst also pursuing a credible evaluation approach to the collection, analysis and reporting of evaluative data.

The political context of evaluation

Evaluation has been depicted as a complex process that has to balance different interests. It is a process that is 'saturated with political concerns' (Berk & Rossi 1990, p. 13). Taylor & Balloch (2005, p. 1) suggest that evaluation is 'socially constructed and politically articulated'. The presence of differing and sometimes competing interests amongst stakeholder groups in evaluation has been recognised as a distinguishing feature of the discipline (Alkin, Hofstetter & Xiaoxia 1997; Berk & Rossi 1990; Guba & Lincoln 1989a; House 1993; Patton 1997). It would be naive to suggest that evaluation can operate in such a highly politicised and interests-driven environment without being significantly influenced or affected by it.

Political dimensions to evaluations commence from the time it is decided to evaluate, as decisions are made about the purpose and role of the evaluation (Taylor & Balloch 2005, p. 8). There are subsequent political decisions made throughout the evaluation about membership, budget, timelines, scope, focus, boundaries, design, method, detail, findings, recommendations and dissemination. In relation to all these aspects of the evaluation, the evaluator can become involved in complex discussions and negotiations about what is considered necessary for a credible evaluation and what is politically desirable or pragmatically acceptable.

Evaluations are often commissioned by government departments and organisations with particular policies and programs they need to implement. Commissioners and stakeholders may desire certain approaches or outcomes from an evaluation that intersect with their organisational mandate and interests. They want programs to continue, to alter in certain directions, or to cease, and anticipate evaluations supporting their respective interests in. Pressure can be placed on the evaluator to identify certain findings from the evaluation and make specific recommendations (positive or negative depending on the circumstance) that are consistent with prevailing political agendas. Policy shifts over the course of the evaluation frequently result in changes to the expectations of the evaluation process, its focus and its results. Pressures emanating from the political context present challenges for evaluators in maintaining the degree of independence and objectivity required during the evaluation process and when interpreting findings. Personal pressures on the evaluator can result in them accommodating political agendas and interests with a potential decline in the quality and level of trustworthiness of the evaluation.

The response of the evaluator to the influences brought to bear by political interests and policy agendas can be varied. Commercial providers of evaluation services can struggle to balance their intent to maintain objectivity and independence with their business interests. For example, an evaluation consultant may find that a program is not working well when assessed against its purpose and objectives and identify findings leading to recommendations to guide program redesign or improved performance. It may not be in the interests of the commissioner of the evaluation to have the program portrayed as under-performing. The commissioner of the evaluation may indicate to the evaluation consultant that the data requires reinterpretation and findings need to be softened or moderated. The evaluation consultant may thus experience a conflict of interest between meeting the needs of the client (and potentially gaining further contracts) and maintaining the objectivity and independence of their findings.

Maintaining objectivity and independence can also be a challenge for the internal evaluator, as they have to balance independence with organisational imperatives and loyalties, and their own career progression. For example, an internal evaluator may experience the same dilemma as depicted above, but risk losing support within their organisation. This dilemma could be heightened if the commissioner of the evaluation is the immediate line manager of the internal evaluator.

The evaluator can thus feel compelled to produce findings or outcomes that are consistent with those anticipated from the commissioning client and in line with their broader political, economic and organisational imperatives. Pressures can be exerted on the evaluator to highlight positive feedback

and play down negative or critical feedback for a program, or alternatively to highlight negative feedback and reduce the level of positive feedback for a program the commissioner of the evaluation is hoping to scale back or close. There can also be pressures to ensure that findings are palatable to the service system or community, in order not to 'rock the boat' in delicate interface relationships between government, service providers, communities and beneficiaries.

Exemplar case studies

The following case studies are exemplars based on real situations faced by the author that illustrate the challenges that can confront an external evaluator working in a politically influenced context.

In the case studies below, the evaluators were placed under considerable pressure to review the data and findings emerging from the evaluation process either to portray the program more positively, or to downplay achievements. In both scenarios, the pressures placed on the commissioners of the evaluations related to prevailing government agendas that influenced their approach and responses. In Case Study 1 such pressures emanated from the political process taking place in advance of an election and in Case Study 2 to a contraction in the mandate of the department and the removal of programs that did not fit the new purpose. The tensions between the commissioners and evaluators emerged most strikingly at the point of writing the first draft of the final report, inclusive of recommendations. In both instances, the evaluations were not well synchronised with the decision-making process, with Case Study 1 making refunding decisions before the evaluation was completed, and in Case Study 2 the program being terminated mid-way during the evaluation.

Although the evaluators were prepared and willing to receive feedback from the commissioning clients (e.g. when there had been errors in the facts presented, missing data or information, questionable analyses, or a need for greater clarity in expression and meaning), the feedback provided by the commissioning clients proceeded down the path of overt pressure to change the content of the findings and the construction of the recommendations.

CASE STUDIES

Case study 1

The evaluation was commissioned by a state government department and commenced 12 months prior to the completion of the pilot period of three-year funding provided for this program. One of the aims of the evaluation was to determine if ongoing funding should be provided to continue the initiative. The evaluation found the program to be underperforming in several respects according to its stated goal and objectives and so a number of recommendations for program redesign were made. The findings emerging from the evaluation were relayed to the commissioners of the evaluation as soon as they became evident.

Towards the end of the 12-month evaluation period a state election was announced. The political party in government at the time was concerned with the particular social issue that this program was established to address. The commissioning client expressed their imperative to showcase successful strategies that the government had funded on this particular social issue. The decision had been made by the state government to re-fund the initiative for a further three-year period with no substantial alteration to the design of the model or its method of delivery.

The evaluators were placed under pressure to review and reframe the data, rephrase the findings and reword the recommendations, in order to provide an overall, more positive evaluation than the data supported. The commissioning client attempted to amend the draft report to indicate where such changes to the text should be made.

The situation was resolved by the evaluation team largely adhering to the contents of the original text used in the final report with minor modifications. It is not known how the evaluation report was subsequently used or applied, or if it was 'shelved'. It is suspected that the latter was the most likely outcome.

Case study 2

During the conduct of an evaluation of a program it became clear that the commissioning client had decided to cease its funding. Program staff members were given notice and midway during the evaluation the program ceased to operate.

The methodology for the evaluation generated data to indicate that the program had been performing well according to its goal and objectives and that it had developed a great level of support from the target communities. This data was assembled into a final evaluation report.

During discussions over the contents of the draft final report pressure was placed on the consultants to adopt a more critical response to the program and to identify greater areas of under-performance than the data supported.

During the final presentation of the report it was suggested by the commissioner that a greater focus should have been placed in considering the appropriateness of the program model for auspice by the particular government department rather than its overall effectiveness and results.

Pawson (2006) supports such experiences of the intersection between policy and evaluation practice emerging from these case studies by commenting on his own experiences with evaluations that had been tossed around as a consequence of changes in political winds. He describes how 'the policy axe had fallen before the research had even gained its stride' (p. 174). Pawson also comments that 'politics, in the last analysis will always trump research' (p. 175).

Thus there are dilemmas for the evaluator when they face a distinct choice between adhering to an objective process or accommodating the organisational imperatives of the commissioning client. There are also challenges for purchasers of evaluation services in using the services of an external or internal evaluator that brings with them objectivity and independence through use of an evidence-based process of data gathering, assessment, analysis and findings and recommendations. It is suggested that greater levels of knowledge and skills are required on behalf of both evaluators and commissioners to negotiate the complex interconnections between politics and evaluation.

The stakeholder context in evaluation

The importance of stakeholders and their participation in evaluation has been recognised by many evaluation theorists. There are different rationales about why stakeholders are considered essential to the process of evaluation. Two prominent aspects of current evaluation practice and theory are the concepts of utilisation and participation. The focus on the aim of increased utilisation of evaluation findings, together with the increasing value placed on participation as part of an empowerment approach to evaluation, have placed more emphasis on the involvement of stakeholders in the evaluative process. Stakeholders are defined as those who have a stake in the evaluation process or a vested interest in the outcome of the evaluation. Stakeholders can include policymakers, funders, program planners, managers, program deliverers, practitioners, community members and program beneficiaries (Guba & Lincoln 1981, 1989a; Weiss 1983a, 1983b). There have been a range of models of stakeholder involvement developed based on either utilisation or participation principles, or a combination of both. Examples of models primarily based on principles of utilisation include Stake's responsive model (1983); Patton's utilisation focused evaluation (1997); Guba and Lincoln's fourth-generation model (1989b); and Byrk's stakeholder-based-evaluation (1983). Cousins and Earl (1995) also proposed a model of participatory evaluation as a means of increasing the relevance of social inquiry knowledge for the benefits of organisational learning and change. Fetterman, Kaftarian and Wandersman (1996) meanwhile, put forward a model of empowerment evaluation, which emphasised self-determination.

Participatory evaluation models are not defined through their use of common methodologies, but rather through their underpinning values. These values include shared ownership of the process, empowerment of program deliverers and beneficiaries to become active participants in the process, collaborative learning, and decision-making. Participatory models thus transcend the use of evaluation solely for accountability to funders by supporting a process that builds the capacity of program participants and facilitates shared learning (Dugan 1996; Estrella 2000; Mikkelsen 2005). The degree of participation embedded in the evaluation, along the continuum of high to low, determines: how collaborative the process is in identifying what will be monitored and evaluated; how and when data will be collected and analysed; how it is interpreted; and how findings will be shared and used to inform decisions.

While there is growing support for participatory approaches to evaluation, there can be a concern with challenges to evaluation rigour through the use of such approaches, due to reduced capacity for objectivity and independence. Lennie (2006) argues that evaluation rigour does not need to be compromised when adopting participatory approaches. She identifies a number of strategies that can be used in participatory evaluations to ensure rigour, such as relationship building with stakeholders, data triangulation, critical reflection, external validation, rigorous data analysis, and stakeholder review of data. However, such strategies do require the involvement of skilled and experienced evaluators with the time and funding available to implement the methods.

Adopting a participatory approach to stakeholder involvement in evaluation thus needs to be a process that is clearly defined in terms of the expectations and boundaries of such involvement. Similar to negotiating the political context of evaluation, the stakeholder context creates a number of challenges for the evaluator and the evaluation project if not carefully planned and managed. Adopting participatory approaches on their own will not be sufficient to reconcile the reality of delivering an evaluation that is both credible and capacity building at the same time.

Credible evaluations

Credibility in evaluation is often described using a range of different terms such as accurate, fair, believable, honest, balanced, defensible, valid, reliable, justifiable, unbiased and impartial. Patton (1997, pp. 260–261) states the credibility and therefore utility of an evaluation are affected by the steps we take to explain our evaluative decisions. In other words, an evaluation has to be defensible to be useful. The concepts of credibility and trustworthiness in evaluation are well described by Schwartz & Mayne (2005, p. 2) who outline their concerns regarding the credibility of evaluative information:

> The success of the current boom in the use of evaluative information will remain largely dependent on its credibility. Program evaluations, performance reports and performance audits all claim to provide objective representations of the reality of program outputs and outcomes, economy, efficiency and effectiveness. Perceptions that evaluative information misrepresents reality (intentionally or not) are likely to render it useless—other than as a tactical weapon in political and bureaucratic skirmishes. There is some evidence suggesting the risk of a credibility crisis regarding much evaluative information.

A number of authors have also supported the perceived threat to the credibility of evaluative information resulting from political and organisational pressures. Schwartz and Mayne (2005, p. 2) state that 'observers of program evaluation practice have long warned that political and commercial pressures on evaluation clients and on evaluators lead to a priori bias in evaluation reports (Chelimsky 1987; Palumbo 1987; Schwartz 1998; Weiss 1973; Wildavsky 1972)' (p. 2). They also go on to suggest that one major threat to the credibility of evaluative information concerns the political, organisational and commercial pressures that can result in bias in evaluation reporting.

Therefore, the challenge for the evaluator is to respond appropriately to the inevitable influences resulting from political and policy considerations, together with the increasing involvement of a range of stakeholders with differing interests.

Independence and objectivity

The terms 'independence' and 'objectivity' are often used interchangeably in evaluation to depict the process of adopting an autonomous and impartial position in the conduct of an evaluation. They are distinct, but inextricably linked concepts. Independence generally refers to the evaluator being awarded the freedom to conduct the evaluation without undue control being exerted by the commissioners of the evaluation, the organisation or program delivery personnel. Objectivity refers to the evaluator's capacity to undertake unbiased and objective assessments and form conclusions during the evaluation.

For the purpose of this article, independence refers to the freedom of the evaluator to pursue the rigour of the evaluation process without compromising the imperatives and pressures emerging from the immediate political and organisational context, the commissioners of the evaluation or its associated stakeholders. Objectivity refers to the impartiality exercised by the evaluator during the selection of evaluation methodology, the approach to the conduct of the evaluation, and the interpretation of findings.

Both internal and external evaluator perspectives are considered here, with the assumption that evaluators working from either of these positions will attempt to approach evaluations with independence and objectivity. Both groups can experience pressures to compromise their independence and their objectivity, whether due to the pursuit of commercial imperatives, or due to organisational loyalties and effects on career advancement prospects.

A range of ethical and practice dilemmas arise for evaluators when they attempt to preserve their independence and objectivity, and these come to light particularly during the stages of identification of findings and presentation of recommendations. Postmodernist theorists would argue that all interpretations of data that result in findings and recommendations are subjective, arising from our personal position, values and orientation in life. Thus, the same set of data can be interpreted in different ways depending on the perspective being used to interpret it. It would be difficult to argue that our subjective lenses do not impact significantly on data analysis and interpretation. Furthermore, Patton (1997) claims that utilisation-focused program evaluation transcends the notion of the pursuit of pure objectivity to attain fairness and balance by placing an emphasis on using appropriate, credible and useful data. Conley-Tyler (2005) also notes the flaws in the adoption of an objectivist approach, and asserts that the best that can be achieved is evaluator impartiality. Taylor and Balloch (2005) describe the paradigm wars taking place in evaluation between those arguing for an independent reality capable of objective description and those arguing that knowledge is contextual and subjective—with the evaluator acting as facilitator and negotiator of different perspectives characterised by multiple views, audiences and accountabilities.

It is my view, based on attendance at many Australasian Evaluation Society conferences, that on balance, contemporary evaluators in the Australasian context are less likely to want to be depicted as evaluation experts using objective scientific research techniques. Rather, they would generally acknowledge that they operate as facilitators of a process of inquiry conducted in a political environment characterised by the involvement of multiple players and perspectives. However, contemporary evaluators, on balance, also place high value on accuracy, impartiality and defensibility in reporting sources of data and in their evaluation findings. They also marshal evidence to support their conclusions. It is this approach of effectively combining a more contemporary policy, influenced by a participatory model of evaluation with evaluation rigour, that requires further exploration and development in the evaluation literature.

The question is how to manage the intersection of contextual influences on the evaluation, with goals of evaluator independence and objectivity, in the best way. Strategies and approaches for operating successfully within the political and stakeholder environment without forfeiting a degree of independence and objectivity in the process are vital for the future of evaluation practice.

Evaluation standards may assist this process. These are considered briefly in the following section.

Evaluation standards

Evaluation societies have developed ethical codes, practice guidelines and standards to guide and inform the conduct of evaluations (AEA, 2002; AES 2000, 2006). Fraser (2001) identifies the key differences between these documents. He states that ethics are about right and wrong whereas standards are about quality and adequacy.

Swartz and Mayne (2005) support the development of standards for evaluation as a first step in quality assurance. These authors analysed evaluation and quality assurance mechanisms in the European Union, the World Bank, and in the four countries of Canada, France, the Netherlands and Switzerland. The authors identified three types of standards, these being product quality, process quality and usefulness. Their investigation of product quality revealed similarity in the standards concerning objectivity including:

> substantiated and impartial/objective findings/ conclusions. The findings and conclusions presented should be supported by the evidence gathered (data and analysis) and should be presented in an impartial (objective) manner (p. 6).

Meaanwhile, the evaluation standards developed by the Swiss Evaluation Society (2000) provide a specific section on what they term 'neutral reporting'. They state one principle regarding evaluator independence, which is that:

> Many different perspectives exist in the environment of evaluation. Stakeholders themselves often hold diverging views of the object of an evaluation. Any given evaluation also runs the danger of being instrumentalized or captured by a particular group or interest, though an evaluation should avoid adopting any one specific point of view. Rather, it should be concerned to represent all relevant interests fairly, and it is important for that reason that an evaluation should take as independent position as possible. An evaluation should avoid being too closely linked to those who have commissioned it, but should also avoid being too close to those persons who are responsible for the object of the evaluation. (Swiss Evaluation Society 2000, p. 12)

The above quotes illustrate some attempts that have been made by evaluation societies to produce evaluation standards that encourage evaluators to maintain their professional independence and objectivity.

Practice standards and guiding principles, ethical guidelines and codes developed by evaluation societies provide some guidance and direction for evaluators in negotiating issues of objectivity and independence, and ideally should be adopted as a framework for the evaluation at the outset. The Australasian Evaluation Society (AES) has developed *Guidelines for the Ethical Conduct of Evaluations* (2006). However, these do not specifically refer to issues of independence or objectivity. Fraser (2001) has argued that the AES should develop professional 'threshold' standards that identify minimum requirements to be met before an evaluation product or process can be judged to be of acceptable quality. There is merit in the further consideration of the recommendation of Fraser (2001) that the Australasian Evaluation Society develop its own evaluation standards to address the broader range of practice issues and dilemmas, and thus provide greater guidance to commissioners and practitioners of evaluation.

Conclusion

The political and stakeholder context of evaluation provides an exciting, if somewhat challenging environment for evaluators to operate within. On the one hand, it is beneficial to see evaluation involving and encouraging the active participation of diverse and varied stakeholders, with the range of interests and mandates they represent. It is also worthwhile placing the evaluation in its policy context to ensure it is relevant, worthwhile and useable. On the other hand, the political and stakeholder context poses some challenges for the evaluator, where interests can result in pressures that impact on the independence and objectivity of the evaluation process. This tension often results in the emergence of a range of ethical and professional dilemmas for the evaluator that requires the evaluator to develop clarity about their roles, responsibilities and boundaries.

This article has argued that the evaluator needs to be proactive about protecting the independence and objectivity of the evaluation process in order to produce a credible evaluation product. The evaluator has to juggle professional conduct that preserves the integrity of the evaluation with values of participation, inclusion and responsiveness to context. This article has also argued that commissioners of evaluations need to develop greater levels of awareness of the value of independence and objectivity in the pursuit of quality and credible evaluations.

There are many levels of potential response to the tensions that emerge in managing the relationship between evaluation and its political environment. Evaluation societies have a role to play both in developing practice and ethical standards for evaluators and in increasing the awareness of commissioners of evaluation about the requirements for, and benefits of, an evaluation that is rigorous, defensible and credible. Large contracting organisations and bodies also have a role to play in specifying codes of conduct for evaluations that they commission.

For the evaluator, an important first step is to outline the requirements for objectivity and independence on the part of the evaluator during initial contract negotiations and the establishment phase of the evaluation, specifying compliance with practice standards or ethical guidelines. The development of practice standards by the Australasian Evaluation Society would assist this process further. Having external and recognised codification to assist negotiations and specifying compliance with ethical codes or practice standards in evaluation contracts could work towards the achievement of shared understandings of the need for evaluations to operate within clear parameters. It may also be appropriate to identify at the commencement of an evaluation, the methods or approaches that could be used if dilemmas or difficulties arise along the way, with clear agreed conflict resolution strategies.

At a more interpersonal level there is reliance on the skills of evaluator as negotiator to manage the evaluation process in order to ensure that the ethical and practice dilemmas that arise are well managed. The role of evaluator as negotiator (Markiewicz 2005) is critical to the success of the evaluation. The evaluator requires skills in managing a process where the political conflicts that underpin an evaluation are exposed and discussed, with strategies developed for their resolution.

Discussion in preceding pages has identified a range of issues that require further consideration in order to improve the interaction between evaluation and its political context and to produce credible evaluations that reflect principles of independence and objectivity. Further work is required to outline in greater detail, the nature of the interface relationships between evaluation and its political context, and the implications of this for evaluation practice. In this regard, it is reassuring to see recent texts that examine the relationship of evaluation to the political and policy context (e.g. Taylor & Balloch 2005; Pawson 2006). This is a journey that both the evaluator and stakeholders involved in evaluations must continue to travel, hopefully armed with greater levels of knowledge and skills about the nature and features of this relationship and the benefits to be achieved by supporting a credible evaluation product.

References

Alkin, M, Hofstetter, C & Xiaoxia, A 1997, 'Stakeholder concepts in program evaluation', in A Reynolds & A Walberg (eds), *Evaluation for Educational Productivity,* JAI Press, Greenwich, Connecticut.

American Evaluation Association (AEA) 2002, *Guiding principles for evaluators*, <http://www.eval.org/Publications/GuidingPrinciples.asp>.

Australasian Evaluation Society (AES) 2000, *Code of ethics*, <http://www.aes.asn.au/about>.

Australasian Evaluation Society (AES) 2006, *Guidelines for the ethical conduct of evaluations*, <http://www.aes.asn.au/about>.

Berk, RA & Rossi, PH 1990, *Thinking about program evaluation,* Sage, Newbury Park, California.

Byrk, AS (ed.) 1983, *Stakeholder-based evaluation. New directions for program evaluation,* no. 17, Jossey-Bass, San Francisco.

Conley-Tyler, M 2005, 'A fundamental choice: internal and external evaluation?, *Evaluation Journal of Australasia*, vol. 4, nos 1 and 2, pp. 3–11.

Cousins, JB & Earl, LM (eds) 1995, *Participatory evaluation in education: studies in evaluation use and organisational learning*, Falmer Press, London.

Dugan, MA 1996, 'Participatory and empowerment evaluation', in DM Fetterman, SJ Kaftarian & A Wandersman (eds), *Empowerment evaluation,* Sage, Thousand Oaks, California.

Estrella, M 2000, *Learning from change: issues and experiences in participatory monitoring and evaluation*, The International Development Research Centre, Ottawa.

Fetterman, DM, Kaftarian, SJ & Wandersman, A (eds) 1996, *Empowerment evaluation,* Sage, Thousand Oaks, California.

Fraser, D 2001, 'Beyond ethics: why we need evaluation standards', *Evaluation Journal of Australasia*, vol. 1, no. 1, March, pp 53–58.

Guba, EG & Lincoln, YS 1981, *Effective evaluation,* Jossey-Bass, San Francisco.

Guba, EG & Lincoln, YS 1989a, *Fourth generation evaluation,* Sage, Newbury Park, California.

Guba, EG & Lincoln, YS 1989b, 'The countenances of fourth-generation evaluation: description, judgment, and negotiation', in DJ Palumbo (ed.), *The politics of program evaluation,* Sage, Newbury Park, California.

House, ER 1993, *Professional evaluation,* Sage, Newbury Park California.

Lennie, J 2006, 'Increasing the rigour and trustworthiness of participatory evaluations: learnings from the field', *Evaluation Journal of Australasia*, vol. 6, no. 1, pp. 27–35.

Markiewicz, A 2005, 'A balancing act: resolving multiple stakeholder interests in program evaluation', *Evaluation Journal of Australasia*, vol. 4, nos 1 and 2, pp 13–21.

Mikkelsen, B 2005, *Methods for development work and research*, 2nd edn, Sage, New Delhi.

Patton, MQ 1997, *Utilisation-focussed evaluation,* 3rd edn, Sage, Thousand Oaks, California.

Pawson, R 2006, *Evidence-based policy: a realist perspective,* Sage, London.

Schwartz, R & Mayne, J 2005, 'Assuring the quality of evaluative information: theory and practice', *Evaluation and Program Planning*, vol. 28, no. 1, pp 1–14.

Stake, R 1983, 'Stakeholder influence in the evaluation of cities-in-schools', in AS Bryk (ed.), *Stakeholder-based evaluation. New directions for program evaluation,* no. 17, Jossey-Bass, San Francisco.

Swiss Evaluation Society 2000, SEVAL Standards, <www.seval.ch/en/documents/SEVAL_Standards_2000_en.pdf>.

Taylor, D & Balloch, S 2005, *The politics of evaluation: participation and policy implementation*, The Policy Press, Bristol, England.

Weiss, C 1983a, 'The stakeholder approach to evaluation: origins and promise', in AS Bryk (ed.), *Stakeholder-based evaluation. New directions for program evaluation*, no. 17, Jossey-Bass, San Francisco.

Weiss, C 1983b, 'Toward the future of stakeholder approaches in evaluation', in AS Bryk (ed.), *Stakeholder-based evaluation. New directions for program evaluation*, no. 17, Jossey-Bass, San Francisco.

REFEREED ARTICLE
Evaluation Journal of Australasia, Vol. 8, No. 2, 2008, pp. 42–48

Program clarification: an overview and resources for evaluability assessment, program theory and program logic

This contribution to the journal examines program clarification for evaluation purposes. It traces the development of this approach over the past three decades, during which the terms evaluability assessment, program theory and program logic have been applied in turn. This is followed by an extensive list of resources that either discuss one of the terms generally or describe applications to a range of program areas.

Rosalind Hurworth

Ros Hurworth is Director of the Centre for Program Evaluation at the University of Melbourne. Email: <r.hurworth@unimelb.edu.au>

The development of clarification approaches

Evaluability assessment (1970s onwards)

Until about 30 years ago, the emphasis in evaluation was on determining impact for accountability purposes. However, evaluators often ran into problems as they tried to achieve this, because programs could have vague or unspecified goals, which made measurement of outcomes well-nigh impossible. Alternatively, programs could be so complex that it was difficult to understand how they worked in practice and so questions arose about program elements and how they could be evaluated.

Attempts to overcome such problems by determining the extent to which a program is ready for evaluation, led to new terminology and to the emergence of writing by evaluators who are now 'household names' for the profession. For example, in the 1970s Joseph Wholey, faced by such difficulties, was credited with devising (and then wrote about) how to overcome them.

To facilitate impact evaluation, he and others worked with program managers and staff to devise models of programs that could reveal program objectives and agreed performance indicators. The work entailed identifying relationships between, and external influences on, program events. Such in-depth examinations led to increased clarity about goals and objectives and identified whether a particular program was coherent, plausible and measurable. As a result of such work, it became possible to determine which program elements were amenable to further evaluation and which were not.

From then on the process enabled evaluators to acquire detailed, firsthand knowledge of programs that could lead to the development of tailored evaluation designs. This descriptive and analytic process became known as Evaluability Assessment (EA), the objective of which was to: 'determine the extent to which a program is ready for evaluation, the changes needed to make the program more manageable and accountable, and toward what questions a more extensive evaluation might usefully be directed' (Schubert 1982, Abstract). In other words, 'EA is a diagnostic and prescriptive technique that can be used to determine the extent to which different problems inhibit program evaluation' (Wholey 1987).

Indeed, Scherzer (2008) goes further to describe EA as a 'pre-evaluation analysis used to used to determine whether program performance is likely to produce desired results and to increase the usefulness of subsequent evaluations'.

Since the 1980s EA has been applied to a wide variety of programs, disciplines and settings (Trevisan 2007) and most commonly has involved collecting information through document reviews, site visits and interviews.

The emergence of program theory (1980s onwards)

Even so the term and ideas behind EA appear to have declined somewhat as the emphasis moved from solely determining ways to evaluate impact to the task of drawing up a 'map' of a program as a task in its own right, that is, defining a detailed theory underpinning a program. This was required because there was a desire to clarify programs for purposes other than determining impact, such as the need to improve program design, ensure better program delivery and to assist in the development of stronger policy. Subsequently, the role of evaluators has increasingly become one of developing and testing program theory founded on the writings of Huey Chen, Carol Weiss and others during the 1980s and 1990s.

Since then, evaluators have increasingly used substantive knowledge to update, simplify, clarify and make more accessible the underlying theory of programs to inform stakeholders. This involves documentation of the assumptions implicit in program design and an indication of the data required to test these assumptions. The process also identifies links between planned activities and anticipated outcomes.

As a consequence of such developments, the application of program theory developed rapidly and moved from being used just in areas such as health promotion to a wide range of program areas such as energy conservation, community-based initiatives and housing (Rogers & Weiss 2007).

Program logic becomes the common term (1990s onwards)

Gradually the term 'program logic' has replaced, or been used synonymously with, program theory over the last two decades. According to Chen, program logic can be defined as 'a set of interrelated assumptions, principles and/or propositions to explain or guide social actions'. Torvatn (1999) calls it 'chains of reasoning' providing a clear framework of the working and functions of a program.

A program logic is usually conveyed visually by diagrams, flow charts or 'trees'. Linney & Wandersman (1991) define such a display as:

> a logical, graphically depicted series of statements that link a problem to the community that it exists in, the possible barriers to solving the problem, the activities and resources that are necessary to address the problem, short-term activities that result from these activities and the hoped for long-term outcomes of the program.

Elements or statements in the model can be determined through methods such as document analysis, concept mapping, interviews or focus groups.

People often see these visual documents as a means to see how theory is linked to implementation and outcomes and also to see if these elements are aligned properly. They are also considered particularly powerful if devised in conjunction with stakeholders. It is a way to clarify underlying assumptions and to reach group consensus. Some also recognise that these charts are not just 'one-off' constructions but are 'living', dynamic documents that can be adapted as environmental and political contexts change.

The contributions of Australasians

At this point we should acknowledge that Australasians have contributed, and added to, ideas about clarification along the way. For example, John Owen has considered the development of what he calls Clarificative Evaluation as a particular form of evaluation (Owen 2006); Patricia Rogers (2000, 2007), along with Jane Davidson (2000), has contributed to ideas about causality in program theory; Rick Cummings and Colin Sharp have considered the application of program theory to educational setting and organisational learning respectively; and Doug Fraser has written about visual displays. Perhaps, though the most influential work has been carried out by Sue Funnell (1997, 2000) who suggested enhancing the usefulness of program theory and logic by developing a matrix where entries have to be made in relation to program contexts, success criteria, potential sources of performance information and criteria for judging such information. This has moved the field on from pure description of the program to providing information that can be used for monitoring purposes.

Further resources

What follows are references that can allow readers to examine the whole area of program clarification further. The material (which is by no means exhaustive) is provided under the major headings of Evaluability Assessment, Program Theory and Program Logic. Each of these sections is then divided in order to present general articles about each term, before showing how the approaches have been applied to particular social science disciplines.

Overall reference

Owen, J 2006, 'Clarificative evaluation', in JM Owen, *Program evaluation forms and approaches*, 3rd edn, Allen and Unwin, Sydney.

Evaluability assessment

General

Ciechalski, JC 1981, 'Planning useful evaluations: evaluability assessment', *Measurement and Evaluation in Guidance*, vol. 13, no. 4, pp. 237–238.

Gantwerk, L 1985, An application and evaluation of evaluability assessment: a case study', *Dissertation Abstracts International*, vol. 46, nos 6–8, p. 2049.

Holloway, WH 1981, *A critique of the evaluability assessment model in organizational analysis*, ED208565.

Jung, SM & Schubert, JG 1983, 'Evaluability assessment: a two-year retrospective', *Educational Evaluation and Policy Analysis*, vol. 5, no. 4, pp. 435–444.

Leviton, LC et al. 1998, 'Teaching evaluation using evaluability assessment, *Evaluation*, vol. 4, no. 4, pp. 389–409.

Nay, JN & Kay, P 1982, *Government oversight and evaluability assessment*, Lexington Books, Lexington, Massachusetts.

Porteous, NL et al. 2002, 'Introducing program teams to logic models: facilitating the learning process', *Canadian Journal of Program Evaluation*, vol. 17, no. 3, pp. 113–141.

Rog, DJ, 1985, 'A methodological analysis of evaluability assessment', *Dissertation Abstracts International*, vol. 46, no. 8B, pp. 2872–2873.

Schmidt, RE 1991, 'Evaluability assessment: a practical approach', *Evaluation and Program Planning*, vol. 14, no. 3, pp. 202–204.

Schmidt, RE et al. 1979, *Evaluability assessment: making public programs work better*, Human services monograph series no. 14, ED183614.

Schubert, JG 1982, *Evaluability assessment: the promise in practice*, ED217078.

Smith, M 1989, *Evaluability assessment: a practical approach*, Kluwer Academic, Boston.

Smith, MF 1989, 'Evaluability assessment', in S Mathison (ed.), *Encyclopaedia of evaluation,* Sage, Thousand Oaks, California.

Smith, NL 1981, 'Evaluability assessment: a retrospective illustration and review', *Educational Evaluation and Policy Analysis*, vol. 3, no. 1, pp. 77–82.

Thurston, WE & Potvin, L 2003, 'Evaluability assessment: a tool for incorporating evaluation in social change programs, *Evaluation*, vol. 9, no. 4, pp. 453–469.

Thurston, WE et al. 2003, 'Evaluability assessment: a catalyst for program change and improvement', *Evaluation and the Health Professions*, vol. 26, no. 2, pp. 206–221.

Trevisan, MS 2007, 'Evaluability assessment from 1986 to 2006, *American Journal of Evaluation*, vol. 28, no. 3, pp. 290–303.

Wholey, JS 1987, 'Evaluability assessment: developing program theory', in L Bickman (ed.), *Using program theory in evaluation. New directions for program evaluation,* no. 33, Jossey-Bass, San Francisco, pp. 77–92.

Wholey, JS 1994, 'Evaluability assessment', *Evaluation News and Comment*, vol. 3, no. 2, pp. 2–13.

Worthington, R 1982, 'Evaluability assessment: making the effort worth the product', *QIER Journal*, vol. 21, pp. 28–34.

Applications of evaluability assessment in particular fields (in alphabetical order of discipline)

Agriculture

Smith, MF 1990, 'Evaluability assessment: reflections on the process', *Evaluation and Program Planning,* vol. 13, no. 4, pp. 359–364.

Community-based programs

Golum, RH 2004, 'An evaluability assessment of a school-based group counselling program for African-American female students living in an urban area of predominantly low-income status families', *Dissertation Abstracts International: Section 8: The Sciences and Engineering*, vol. 64, nos 9–8, p. 4614.

Meeres, SL et al. 1995, 'Evaluability assessment of a community-based program', *The Canadian Journal of Program Evaluation*, vol. 10, no. 1, pp. 103–121.

Piquero, A 1998, 'Applying an evaluability assessment tool to community-based programs in Pittsburgh', *The Prison Journal*, vol. 78, no, 1, p. 74.

Thurston, WE & Ramaliu, A 2005, Evaluability assessment of a survivors of torture program: lessons learned, *Canadian Journal of Program Evaluation*, vol. 20, no. 2, pp. 1–25.

Criminology

Basile, KC et al. 2005, Evaluability assessment of the rape prevention and education program, *Journal of Women's Health*, vol. 14, no. 3, pp. 201–207.

Finckenauer, JO et al. 2005, 'Evaluability assessment in juvenile justice: a case example, *Youth Violence and Juvenile Justice*, vol. 3, no. 3, pp. 265–275.

Matthews, B. et al. 2001, 'Making the next step: evaluability assessment to improve correctional programming, *Prison Journal*, vol. 81, no. 4, pp. 454–472.

Education

Kraetzer, AV & Schofield, RG 1982, *School social work services evaluability assessment*, ED221798.

Quint, J 2001, *An evaluability assessment of the Toyota Families in Schools Program*, *ED465034*.

Roscoe, FP 1987, 'An evaluability assessment of an alternative education program, *Dissertation Abstracts International,* vol. 48, no. 1B, p. 253.

Russ-Eft, D 1986, 'Evaluability assessment of the adult education program: the results and their use', *Evaluation and Program Planning*, vol. 9, no. 1, pp. 39–47.

Scherzer, M 2008, 'An evaluability assesment of a preschool program in a public school for children on the autistic spectrum', *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, vol. 68, no. 9A, p. 3735.

Schofield, RG & Kraetzer, AV 1988, 'An evaluability assessment of school social work services: a tool for social workers and program managers', *Journal of School Social Work*, vol. 2, no. 2, pp. 15–30.

Scott, AC & Bourexis, PS 1982, *Evaluability assessment of the Title 11 Basic Skills Improvement Program: implications for state-level programs*, ED217082.

Spielberger, J & Govette, P 2006, *The Early Childhood Cluster Initiative of Palm Beach County, Florida: early implementation study and evaluability assessment, final report Florida*, Chapin Hall Working Paper.

Verma, S & Marveske, G, *Using evaluability assessment to improve programs in the cooperative extension system*, ED329699.

Yampolsky, PH 1984, 'An evaluability assessment of a special educational–vocational program for handicapped high school pupils', *Dissertation Abstracts International*, vol. 45, nos 2–8, p. 659.

Family programs

Barnow, BS & Stapeton, DC 1997, *An evaluability assessment of responsible fatherhood programs*, ED463838.

Lee, WG 2007, 'An evaluability assessment of a homeless program for families', *Dissertation Abstracts International: Section 8: The Sciences and Engineering*, vol. 67, nos 7–8, p. 4159.

Miller, EN, 2001, 'An evaluability assessment of a community living program serving single parents with histories of homelessness, mental illness and often co-occurring substance abuse and who are living with and parenting their children, *Dissertation Abstracts International: Section 8: The Sciences and Engineering*, vol. 62, nos 7–8, p. 3385.

Health

Casebeer, A & Thurston, WE 1995, 'Evaluability assessment in health care: an example of the patient care and outcome process', *The Canadian Journal of Program Evaluation*, vol. 10, no. 1, pp. 89–102.

Durham, J et al. 2007, 'An evaluability assessment of a nutrition promotion project for newly arrived refugees', *Health Promotion Journal of Australia*, vol. 18, no. 1, pp. 43–49.

Dwyer, J. et al. 2003, 'Maximising children's physical activity: an evaluability assessment to plan a community-based, multi-strategy approach, *Health Promotion International*, vol. 18, no. 3, pp. 199–208.

Macaskill, L et al. 2000, 'An evaluability assessment to develop a restaurant health promotion program in Canada, *Health Promotion International*, vol. 15, no. 1, pp. 57–69.

Strosberg, MA 1983, 'Evaluability assessment: from theory to practice in the Department of Health and Human Services, *Public Administration Review*, vol. 43, no. 1, pp. 66–71.

Mental health

Fisher, EJ & Peters, L 1985, 'The role of evaluability assessment in mental health program evaluation', *Canadian Journal of Community Mental Health*, vol. 4, no. 2, pp. 25–34.

Johncox, V 2000, 'Evaluability assessment of staff training in special care units or persons with dementia: strategic issues', *Canadian Journal of Program Evaluation*, Special issue, pp. 53–66.

Technology

Youtie, J et al. 1999, 'Using an evaluability assessmnet to select methods for evaluating state technology development programs', *Evaluation and Program Planning*, vol, 22, no. 1, pp. 55–64.

Program theory

General

Bickman, L (ed.) 1987, *Using program theory in evaluation. New directions for program evaluation,* no. 33, Jossey Bass, San Francisco.

Birckmayer, J & Weiss, C 2000, 'Theory-based evaluation in practice: what do we learn?', *Evaluation Review*, vol. 24, no. 4, pp. 407–431.

Chen, HT 1989, 'The conceptual framework of the theory-driven perspective', *Evaluation and Program Planning*, vol. 12, no. 4, pp. 391–316.

Chen, HT 1990, *Theory-driven evaluations*, Sage, Newbury Park, California.

Chen, HT & Rossi, P 1983, 'Evaluating with sense: the theory-driven approach', *Evaluation Review*, vol. 7, no. 3, pp. 283–302.

Cook, TD 2000, 'The false choice between theory-based evaluation and experimentation', in PJ Rogers et al. (eds), *Program theory in evaluation: challenges and opportunities. New directions for evaluation,* no. 87, Jossey-Bass, San Francisco, pp. 27–34.

Coryn, CLS 2008, 'Program theory-driven evaluation science: strategies and applications, *American Journal of Evaluation*, vol. 29, no. 2, pp. 215–217.

Davidson, EJ 2000, 'Ascertaining causality in theory-based evaluation', in PJ Rogers et al. (eds), *Program theory in evaluation: challenges and opportunities. New directions for evaluation,* no. 87, Jossey-Bass, San Francisco, pp. 17–26.

Donaldson, SI 2005, 'Using program-theory-driven evaluation science to crack the Da Vinci code', in MC Alkin & CA Christie (eds), *Theorists' models in action. New directions for evaluation,* no. 106, Jossey-Bass, San Francisco, pp. 65–84.

Fear, W 2007, 'Programme evaluation theory: the next step toward a synthesis of logic models and organisational theory. *Journal of MultiDisciplinary Evaluation*, vol. 4, no. 7, pp. 13–15.

Funnell, SC 2000, 'Developing and using a program theory matrix for program evaluation and performance monitoring', in PJ Rogers et al. (eds), *Program theory in evaluation: challenges and opportunities. New directions for evaluation,* no. 87, Jossey-Bass, San Francisco, pp. 91–101.

Leew, FL 2003, 'Reconstructing program theories: methods available and problems to be solved', *American Journal of Evaluation*, vol. 24, no. 1, pp. 5-20.

Patton, MQ 1997, 'Developing program theory', *Evaluation News and Comment*, vol. 6, no. 1, pp. 18-20.

Rogers, PJ 2000, 'Causal models in program theory evaluation', in PJ Rogers et al. (eds), *Program theory in evaluation: challenges and opportunities. New directions for evaluation,* no. 87, Jossey-Bass, San Francisco, pp. 47–55.

Rogers, PJ. et al. 2000, 'Program theory evaluation: practice, promise, and problems', in PJ Rogers et al. (eds), *Program theory in evaluation: challenges and opportunities. New directions for evaluation,* no. 87, Jossey-Bass, San Francisco, pp. 5–13.

Rogers, PJ & Weiss, CH 2007, 'Theory-based evaluation: reflections ten years on' and 'Theory based evaluation: past, present and future', in S Mathison (ed.), *Enduring issues in evaluation: the 20th anniversary of the collaboration between NDE and AEA. New directions for evaluation*, no. 114, Jossey-Bass, San Francisco, pp. 63–81.

Weiss, CH 1997a, 'Theory-based evaluation: past, present, and future', in DJ Rog & D Fournier (eds), *Progress and future directions in evaluation: perspectives on theory, practice, and methods. New directions for evaluation*, no. 76, Jossey-Bass, San Francisco, pp. 41–55.

Weiss, CH 1997b, 'How can theory-based evaluation make greater headway?, *Evaluation Review*, vol. 21, no. 4, pp. 501–524.

Applications of the use of program theory in particular fields (in alphabetical order of discipline)

Community/welfare

Carvalho, S & White, H 2004, 'Theory-based evaluation: the case of social funds', *American Journal of Evaluation*, vol. 25, no. 2, pp. 141–160.

Mercier, C et al. 2000, 'An application of theory-driven evaluation to a drop-in youth center', *Evaluation Review*, vol. 24, no. 1, pp. 73–91.

Solomon, B 2002, 'Accountability in public child welfare: linking program theory, program specification and program evaluation', *Child and Youth Services Review*, vol. 24, nos 6–7, pp. 385–407.

Criminology

Tilley, N 2004, 'Applying theory-driven evaluation to the British Crime Re-education Programme', *Criminal Justice*, vol. 4, no. 3, pp. 255–276.

Education

Crew, RE Jr & Anderson, MR 2003, 'Accountability and performance in charter schools in Florida: a theory-based evaluation', *American Journal of Evaluation*, vol. 24, no. 2, pp. 189–212.

Cummings, R, Stephenson, K & Hale, L 2001, 'Using program theory in an educational setting', *Evaluation Journal of Australasia*, vol. 1, no. 1 (new series), pp. 29–39.

Heubner, TA 2000, 'Theory-based evaluation: gaining a shared understanding between school staff and educators', in PJ Rogers et al. (eds), *Program theory in evaluation: challenges and opportunities. New directions for evaluation,* no. 87, Jossey-Bass, San Francisco, pp. 79–89.

Turnbull, B 2002, 'Program theory building: a strategy for deriving cumulative evaluation knowledge', *American Journal of Evaluation*, vol. 23, no. 3, pp. 275–290.

Health

Bay, C 2005, 'Theory-driven evaluation of nursing care', *Journal of Nursing Scholarship,* vol. 37, no. 1, p. 3.

Frosch, D. et al. 2008, 'Using decision aids in community-based primary care: a theory-driven evaluation with ethnically diverse patients', *Journal of General Internal Medicine*, vol. 23, Supplement 2, p. 434.

Walshe, K 2007, 'Understanding what works and why in quality improvement: the need for theory-driven evaluation', *International Journal for Quality in Health Care*, vol. 19, no, 2, pp. 57–59.

Industry

Torvatn, H 1999, 'Using program theory models in evaluation of industrial modernization programs: three case studies', *Evaluation and Program Planning*, vol. 22, no. 1, pp. 73–82.

Program logic

General

Alter, C & Murty, S 1997, 'Logic modelling: a tool for teaching practice evaluation', *Journal of Social Work Education*, vol. 33, no. 1, pp. 103–108.

Bott, TF & Eisenhawer, SW 1989, 'Programme planning with logic trees', *International Journal of Quality and Reliability Management*, vol. 6, no. 2, pp. 15–24.

Cadwallader, T & Klutz, M, n.d., *Logic model worksheet*, University of Wisconsin, <http://www.uwex.edu/ces/cty/marathon/ag/documents/IncLogicModelFinal.pdf>.

Cooksy, L, Gill, P & Kelly, P 2001, 'The program logic model as an integrative framework for a multimethod evaluation', *Evaluation and Program Planning*, vol. 24, no. 2, pp. 119–128.

Den Heyer, M 2002, 'The temporal logic model concept', *The Canadian Journal of Program Evaluation*, vol. 17, no. 2, pp. 27–48.

English, B & Kaleveld, L 2003, 'The politics of program logic', *Evaluation Journal of Australasia,* vol. 3, no. 1, pp. 35–42.

Fashola, O 2001, 'Logic model basics', *Harvard Family Research Project*, vol. 7, no. 2, pp. 14–15.

Fraser, D 2001, 'Visualising program logic: two new graphic conventions', *Evaluation Journal of Australasia,* vol. 1, no. 2, pp. 54–60.

Frechtling, JA 2007, *Logic modelling methods in program evaluation*, Jossey Bass, San Francisco.

Funnell, S 1997, 'Program logic: an adaptable tool for designing and evaluating programs', *Evaluation News and Comment*, vol. 6, no. 1, pp. 5–7.

Ganley, HE & Ward, M 2001, Program logic: a planning and evaluation method, *Journal of Nursing Administration*, vol. 31, no. 1, pp. 4, 39.

Gasper, D 2000, Evaluating the 'logical framework approach' towards learning-oriented development evaluation', *Public Administration and Development*, vol. 20, no. 1, pp. 17–28.

Gill, JR et al. 1998, 'Program logic model: a tool for evaluating social change', paper presented at the American Evaluation Association Annual Conference, Chicago.

Hernandez, M 2000, 'Using logic models and program theory to build outcome accountability', *Education and Treatment of Children*, vol. 23, no. 1, pp. 24–40.

Ince, R 1994, 'DIY evaluation and program logic', *Evaluation Journal of Australasia*, vol. 6, no. 8, pp. 56–60.

Israel, GD 2001, *Using logic models for program development*, University of Florida, IFAS Extension, <http://edis.ifas.ufl.edu/pdffiles/WC/WC04100.pdf>.

Julian, D 1997, 'The utilization of the logic model as a system level planning and evaluation device', *Evaluation and Program Planning*, vol. 20, no. 3, pp. 251–257.

Julian, DA, Jones, A & Deyo, D 1995, 'Open systems evaluation and the logic model: program planning and evaluation tools', *Evaluation and Program Planning*, vol. 18, no. 4, pp. 333–341.

Kirkpatrick, S 2001, 'The program logic model: what, why and how?', *Ontario health promotion e-mail bulletin,* <http://www.charityvillage.com/cv/research/rstrat3.html>.

Lenne, B 1987, 'Describing program logic', *Program Evaluation Bulletin of the NSW Public Service Board*, Sydney.

McLaughlin, JA & Jordan, GB 1999, 'Logic models: a tool for telling your program's performance story', *Evaluation and Program Planning*, vol. 22, no. 1, pp. 65–72.

Renger, R & Titcomb, A 2002, A three-step approach to teaching logic models', *American Journal of Evaluation*, vol. 23, no. 4, pp. 493–503.

Rush, B & Ogbourne, A 1991, 'Program logic models: expanding their role and structure for program planning and evaluation', *Canadian Journal of Program Evaluation*, vol. 6, no. 2, pp. 95–106.

Sartorius, R 1991, 'The logical framework approach to project design and management', *Evaluation Practice*, vol. 12, no. 2, pp. 139–147.

Taut, S 2008, 'Logic modelling methods in program evaluation', *American Journal of Evaluation*, vol. 29, no. 2, pp. 217–218.

Taylor-Powell, E 1998, *The logic model: A program performance framework,* University of Wisconsin—Extension, Madison, Wisconsin.

Torghele, K. et al. 2007, 'Logic model use in developing a survey instrument for program evaluation', *Public Health Nursing*, vol. 24, no. 5, pp. 472–479.

Torvatn, H 2008, 'Logic modeling methods in program evaluation', book review, *Evaluation and Program Planning,* vol. 31, no. 2, pp. 219–221.

WK Kellogg Foundation 2004, *Logic model development guide*, WK Kellogg Foundation, Battle Creek, Michigan.

Wyatt Knowlton, L & Phillips, CC 2008, *The logic model guidebook: better strategies for great results*, Sage, Thousand Oaks, California.

Methods to construct program logics

Anderson, LA et al. 2006, 'Using concept mapping to develop a logic model for the prevention research centre's program', *Preventing Chronic Disease*, vol. 3, no. 1, p. A06.

Guigu, PC & Rodriguez-Compos, L 2007, 'Semi-structured interview protocol for constructing logic models', *Evaluation and Program Planning*, vol. 30, no. 4, pp. 339–350.

Unrau, YA 2001, 'Using client exit interviews to illuminate outcomes in program logic models: a case example', *Evaluation and Program Planning*, vol. 24, no. 4, pp. 353–361.

Yampolskaya, S et al. 2004. 'Using concept mapping to develop a logic model and articulate a program theory: a case example', *American Journal of Evaluation,* vol. 25, no. 2, pp. 191–207.

Applications of program logic to a variety of disciplines in particular fields (in alphabetical order of discipline)

Agriculture

Framst, G 1995, 'Application of program logic models to agricultural technology transfer programs', *Canadian Journal of Program Evaluation*, vol. 10, no. 2, pp. 121–132.

Community/welfare

Chen, WW et al. 1998–9, 'Using a logic model to plan and evaluate a community intervention program: a case study', *International Quarterly of Community Health Education*, vol. 18, no. 4, pp. 449–458.

Edwards, DE 1995, 'A community approach for Native American drug and alcohol prevention programs: a logic model framework', *Alcoholism Treatment Quarterly*, vol. 13, no. 2, pp. 43–62.

Isaacs, B, & Perlman, N, 2004, 'Evaluation of a treatment foster care program: development of program logic', paper submitted for symposium at International Association for the Scientific Study of Intellectual Disability World Congress, 14–19 June, Montpellier, France.

Kolasa, K & Lackey, C 2006, 'The logic model as a framework for community program evaluations: the Food Literacy Partners Program', *Family Medicine*, vol. 38, no. 10, pp. 690–691.

Pathman, D et al. 2003, 'Use of program logic models in the Southern Rural Access Program', *Journal of Rural Health*, vol. 19, no. 5, pp. 308–313.

Rawl, R et al. 2006, 'The logic model as a framework for community program evaluations: the food literacy partners program', *Family Medicine*, vol. 38, no. 10, pp. 690–691.

Criminology

Bureau of Justice Assistance, *Planning the evaluation: developing and working with program logic models*, <http://www.ojp.usdoj.gov/BJA/evaluation/guide/pe4.htm>.

Education

Coffman, J 1999, 'Learning from logic models: an example of a family/school partnership program', *Harvard Family Research Project*, <http://www.gse.harvard.edu/hfrp/pubs/onlinepubs/rrb/learning.html>.

Marcinkowski, T 2004, *Using a logic model to review and analyse an environmental education program*, North American Association for Environmental Education, Washington DC.

Health

Armstrong, EG & Barsion, SJ 2006, 'Using an outcomes-logic-model approach to evaluate a faculty development program for medical education', *Academic Medicine*, vol. 81, no. 5, pp. 483–488.

Cox, RJ 2000, 'Using program logic models in evaluation: a review of the literature and the spinal outreach team experience', *Canadian Journal of Program Evaluation*, vol. 8, no. 1, pp. 117–134.

Dwyer, J 1996, 'Applying a program logic model in program planning and evaluation', *Public Health and Epidemiology Report Ontario*, vol. 7, no. 2, pp. 38-46.

Dwyer, JJ & Makin, S 1997, 'Using a program logic model that focuses on performance measurement to develop a program', *Canadian Journal of Public Health*, vol. 88, no. 6, pp. 421–425.

Dykeman, M. et al. 2003, 'Development of a program logic model to measure the processes and outcomes of a nurse-managed community health clinic', *Journal of Professional Nursing*, vol. 19, no. 4, pp. 197–203.

Fraser, B & Hllett, RG 2003, 'Use of a program logic model to guide the development of a strategic plan for Wellington County Hospitals Network', *Healthcare Management Forum*, vol. 16, no. 3, pp. 12–17.

Hulton, LJ 2007, 'An evaluation of a school-based teenage pregnancy prevention program using a logic model framework', *The Journal of School Nursing*, vol. 23, no. 2, pp. 104–110.

Linney, JA & Wandersman, A 1991, *Prevention Plus 111: assessing alcohol and other drug prevention programs: a four-step guide to useful program assessment*, United States Department of Heath and Human Services, Rockville, Maryland.

McEwan, KL & Bigelow, DA 1997, 'Using a logic model to focus health services on population health goals', *Canadian Journal of Program Evaluation*, vol. 12, no. 1, pp. 167–174.

Moyer, A, Verhovsek, H et al. 1997, 'Facilitating the shift to population-based public health programs: innovation through the use of framework and logic model tools', *Canadian Journal of Public Health*, vol. 88, no. 2, pp. 95–98.

Sitaker, M. et al. 2008, 'Adapting logic models over time: the Washington State Heart Disease and Stroke Prevention Program experience', *Preventing Chronic Disease*, vol. 5, no. 2, p. A60.

Stinson, S & Wilkinson, C 2004, 'Creating a successful clinical extern program using a program planning logic model', *Journal for Nurses in Staff Development*, vol. 20, no. 3, pp. 140–144.

Mental health

Coughlin, P et al. 2008, 'Program logic model: a valued tool for leaders and clinical team members in a multidisciplinary Canadian early intervention in psychosis program', *Early Intervention in Psychiatry*, vol. 2, Supplement 1, p. A112.

Organisations

Sharp, CA & Stock, H 2005, 'In search of a program logic for the evaluation of corporate governance and organisational performance', *Evaluation Journal of Australasia*, vol. 5, no. 2, pp. 48–59.

Social work

Alter, C & Egan, M 1997, 'Logic modelling: a tool for teaching critical thinking in social work practice', *Journal of Social Work Education*, vol. 33, no. 1, pp. 85–103.

Youth

Hermann, JA 1996, 'Using logic models to strengthen service program development: an example of antipoverty programs for at-risk youth', paper presented at the Sociological Practice Association Annual Meeting, Arlington, Virginia.

Jacobs-Lowery, RL 2002, 'A formative evaluation of an adolescent development program: the logic model process', *Dissertation Abstracts International: Section 8: The Sciences and Engineering*, vol. 62, nos 7–8, p. 3379.

2008 AES Awards for Excellence in Evaluation

awards

Peter Bycroft, *Chair, AES Awards for Excellence in Evaluation Committee*

The aim of the AES Awards for Excellence in Evaluation is to encourage the development of evaluation and high-quality evaluation practice in Australasia. In 2008, seven awards were offered by the Australasian Evaluation Society.

Two of the awards recognise individual evaluators nominated by their peers: the Outstanding Contribution to Evaluation (ET&S) Award and the Emerging New Talent Award. Five awards recognise organisations, project teams, specific evaluation projects, programs or systems nominated by the evaluators and/or the organisations involved: the Best Evaluation Publication Award (the Caulley Tulloch Award), the Best Evaluation Study Award, the Best Evaluation Policy and Systems Award, the Community Development Evaluation Award and the Indigenous Evaluation Award.

The following awards were presented at the 2008 AES International Conference dinner in Perth in September.

The Best Evaluation Study Award 2008

This is awarded to an individual or team that has conducted an evaluation study that has made, or has the potential to make, a significant contribution to the practice or use of evaluation in either the public or private sector in Australasia.

The AES Awards for Excellence in Evaluation Committee awarded Simon Smith, Julie McGeary, the Victorian Department of Primary Industries and Dr Martin Andrew, Lili Pechey, Dr Don Burnside and Dr Todd Richie from URS Australia for the *Our Rural Landscape Program Evaluation*.

This nomination covers a complex and multi-level program that demanded a multi-level evaluation. The evaluation design and management were well considered and developed. Stakeholder involvement was a large component, and necessary for the evaluation to get traction. The strength of the nomination is around evaluation management aspects: the many activities and efforts required to keep the momentum of the evaluation, the ongoing communication of the evaluation findings, the client organisation harnessing the evaluation learnings, in terms of administering such a large program, and how to evaluate complex programs.

Three of the winners of the Best Evaluation Study Award 2008 (from left to right): Don Burnside (URS, Perth), Julie McGeary (Victorian Department of Primary Industries, Ballarat) and Martin Andrew (URS, Adelaide) with AES President Jenny Neale.

The Best Evaluation Policy and Systems Award 2008

Kerrie Ikin accepting the 2008 award for Best Evaluation Policy and Systems on behalf of the NSW Department of Education and Training, Western Sydney Region, with AES President Jenny Neale.

This is awarded to an individual or team that has the best integrated design and implementation of evaluation systems or frameworks. Nominees must include a partnership with their clients who will receive a separately titled award (National Evaluation Advocacy Award). The scope includes programs, products, services, personnel, management strategies and other initiatives.

This was awarded to NSW Department of Education and Training, Western Sydney Region, for the *School Cyclical Review Framework*.

This nomination is definitely an example of sound evaluation policy and system, based on research and solid design around education review principles. It is a very well-documented project and gives a very impressive account of working with various stakeholders; getting the principals involved, participating and ready to continue in a sustained way; and building the evaluation capacity within the schools. The judges felt that this was a particularly solid piece of collaborative work that was evidence-based, had the stakeholder engaged, built client understanding of evaluation, and was effective in that the evaluation system is being rolled out across the region.

The Indigenous Evaluation Award 2008

Award winners who attended the 2008 AES International Conference dinner (from left to right): Julie McGeary, Don Burnside, Kerrie Ikin and Martin Andrew, with the recipients of the Indigenous Evaluation Award 2008, Anne Markiewicz (Anne Markiewicz and Associates) and Colin Plowman (Department of Finance and Deregulation).

The aim of the award is to recognise policy, project or program evaluations where the evaluation demonstrates sensitivity, professionalism, excellence and achievement in involving Indigenous communities, addressing Indigenous issues or with a significant benefit to Indigenous policy development. Award winners in this category may also demonstrate the development of evaluation modelling, methods or techniques with a focus on the needs and perspectives of Indigenous communities or the Indigenous sector.

This was awarded to the Department of Finance and Deregulation, Office of Evaluation and Audit (Indigenous Programs) with Anne Markiewicz, Director, Anne Markiewicz and Associates Pty Ltd for the *Evaluation of the FaCSIA Family Violence Programs: Family Violence Regional Activities Program—Family Violence Partnership Program (FVRAPFVPP)*.

The judges regarded this as an excellent example of evaluation in the Indigenous evaluation sector. The approach involved defensible modelling, the development of a structured analytical framework, was culturally sensitive and actively engaged Indigenous and remote communities while delivering an effective evaluation project in a short time frame.

Attention to the design and development of a sound evaluative framework and the use of program logic and modelling in the initial stages provided a sound basis for the subsequent evaluation, rigour to data collection and structure to subsequent reporting.

BOOK REVIEWS

Title:	**The Telephone Interviewer's Handbook: How to Conduct Standardized Conversations**
Author:	Patricia A Gwartney
Publisher/year:	Jossey-Bass (a Wiley imprint), 2007
Extent/type:	315 pages, paperback
Price:	A$33.95/NZ$38.99 from Wiley Australia which offers a 15% discount to AES members, phone 1800 777 474 (within Australia), 0800 448 200 (from NZ only), +61 7 3354 8444 (from overseas), email <custservice@johnwiley.com.au>, website <www.johnwiley.com.au>
ISBN:	978-0-7879-8638-4

The book is an excellent resource for evaluators who wish to gather data via telephone interviews to minimise interviewer-related error and avoid interviewer bias.

Book contents

Chapter 1: Introduction

The author Patricia A Gwartney is professor and associate head in the Department of Sociology at the University of Oregon. The handbook has its foundations based on more than 30 years of survey research experience. Gwartney suggests that, 'interviewers are critical to the success of telephone survey projects and organisations' (p. 2).

Chapter 2: Who Conducts Surveys?

While the book has been written by an author focused on its application in the United States of America, with specific references to survey research companies, the book provides useful concepts, frameworks and tools (pp. 29–35, 103–115, 137–139, 144–154, 163–7, 285–287).

In particular, Gwartney provides useful information about survey company employers and employees (questions to ask, expected answers and unwanted answers) (pp. 36–48).

Chapter 3: Survey Professionalism

Chapter 3 provides the principles for: (a) ethics in survey research, (b) codes of ethics, (c) American Federal Government requirements, (d) unethical pseudosurveys and laws protecting work, and (e) research on survey research.

Chapter 4: What to Expect in Telephone Interviewer Training

Chapter 4 is focused on general, project-specific and refresher training for the telephone interviewer.

Processes covered are: general (activities, testing), project-specific (preliminary activities, pre-contact letter and answers to common questions, survey instrument review including computer-assisted telephone interviewing [CATI], scheduling data collection work shifts, avoiding predictable problems), and refresher (prevention of interviewer skill rusting, interviewer errors, and mini-training sessions just prior to data collection sessions).

The salient point from this chapter is that structured conversations should enable data collection from all respondents.

Chapter 5: Calling

Salient points from the chapter:

- carefully review each respondent's call history before dialling the telephone
- three cardinal rules of scheduling callbacks (AM or PM, time zone, week days)
- vary days of the week to call back reluctant respondents.

Chapter 6: Introducing the Standardized Interview

Salient points from the chapter:

- voice management, word enunciation and speech pace
- telephone script
- samples and data collection (refusals, unusual situations).

Chapter 7: Asking Questions in the Standardized Interview

Salient points from the chapter:

- survey question types
- print conventions
- guidelines for asking questions
- demography (race, ethnicity, occupation, industry)
- recording responses
- unusual circumstances
- ending the interview.

Chapter 8: What to Expect in the Survey Workplace

Salient points from the chapter:

- workplace settings
- workplace routines
- communication with the employer
- employment status, pay and benefits
- workplace policies
- workplace supervisors.

Book review criteria

This book review applies Fellow of the Australian Evaluation Society (AES) Colin Sharp's seven-point framework for book reviews:

1 be comprehensive and eclectic
2 offer a heuristic system
3 provide a valid framework
4 provide case examples from a variety of fields
5 be user-friendly
6 be technically competent
7 provide an adequate explanation of data collection and analysis for the uninitiated reader (Sharp, 2007, p. 60).

1 Comprehensive and eclectic

This comprehensive book is focused only on telephone interview processes, rather than the provision of eclectic world views of standardised conversations for telephone interviewing in survey research.

2 Hueristic system

The many tables, examples and forms enhance learning about telephone interviewing to conduct standardised conversations.

3 Valid framework

The book applies traditional and modern professional practice of data collection for survey research. There are limited in-text references about standardised conversations used for telephone interviewing.

4 Case examples from a variety of fields

As the self-titled handbook was written as a reference for conducting standardised telephone interviews, case examples from a variety of fields would have broadened the reader's perspective.

5 User-friendly

The book is student-friendly with contents, appendix (glossary and example forms), notes, bibliography and index.

6 Technically competent

The author conveys technical competence evident form her three decades of professional practice, which is based on robust research methodologies and sound approaches evident in Chapters 5–7.

7 Data collection and analysis

The book provides adequate coverage of telephone interview data collection methods in survey research. It touches on the basic survey process for study design and planning, survey instrument design, sampling, data collection and entry, data analysis and reporting (pp. 10–14). However, it leaves in-depth explanations about research design and data analysis to other authors (Bogdan & Biklen, 1998; Glaser & Strauss, 1967; Hesse-Biber & Levy, 2006; Owen, 1999; Posavac & Carey, 1997; Strauss & Corbin, 1990; Wiersma, 1995; Worthen & Sanders, 1987).

Best features of the book

The main strengths of this publication are its accessibility as a student text and that it provides a useful framework for conducting telephone interviews. Academics in the higher education sector may find the book to be a useful resource as a prescribed textbook for telephone interviewing data collection methods of research methodologies.

The best features of the book are: (a) Chapter 5: 'Calling'—the in-depth information about CATI (pp. 98–116); (b) Chapter 6: 'Introducing the Standardised Interview'—managing your voice (pp. 124–125), survey (p. 140), refusals (pp. 142–168); and (c) Chapter 7: 'Asking Questions in the Standardized Interview'—summary guidelines for asking questions, especially probing, and race, ethnicity, occupation and industry categories (pp. 203–224).

Readers in the evaluation community should find the book a useful resource for conducting standardised conversations via telephone interviewing.

Opportunity for improvement

The book could be improved by changing the format from a pure handbook into an instructional design one, so that each chapter includes:

- content map
- introduction with learning objectives, assessment and references
- conceptual diagram and theoretical framework
- two-column format with wide column for content and narrow column nearest to the page edge for a highlighted box with the numbered learning objective and a title followed by a definition or key points (see example below)
- highlighted section summaries
- highlighted chapter summaries
- topics for discussion
- exercises
- references.

Learning objective 1

Describe a refusal conversation.

Refusals

A refusal conversation involves developing a strategy tailored to each reluctant respondent to convince him or her to answer the survey questions (Gwartney, p. 161).

While the book could be improved by providing additional information about research survey design, there is no real need to do so as this area is covered adequately by other authors. At the end of Chapter 1, additional references could be provided for further reading. In addition, greater use of in-text referencing would provide stronger points of reference for the book.

References

Bogdan, RC & Biklen, SK. 1998, *Qualitative research for education: an introduction to theory and methods,* Allyn and Bacon, Boston.

Glaser, BG & Strauss, A. 1967, *The discovery of grounded theory: strategies for qualitative research*, Aldine, Piscataway, New Jersey.

Hesse-Biber, SN & Levy, PL, 2006, *Emergent methods in social research*, Sage, Thousand Oaks, California.

Owen, J 1999, *Program evaluation: forms and approaches*, 2nd edn, Allen and Unwin, St Leonards, NSW.

Posavac, EJ & Carey, RG 1997, *Program evaluation methods and case studies,* 5th edn, Prentice Hall, Englewood Cliffs, New Jersey.

Sharp, C 2007, 'Book reviews', *Evaluation Journal of Australasia,* vol. 7, no. 1, pp. 59–60.

Strauss, AJ & Corbin, J 1990, *Basics of qualitative research: grounded theory procedures and techniques,* Sage, Newbury Park, California.

Wiersma, W 1995, *Research methods in education: an introduction,* 6th edn, Allyn and Bacon, Boston.

Worthen, BR & Sanders, JR 1987, *Educational evaluation: alternative approaches and practical guidelines*, Longman, New York.

Reviewed by:

Gloria Carter

Managing Director
Third Generation People
PO Box 1333
Robina, Queensland 4226
Email: <gia2004@bigpond.net.au>

Title:	**Level Best: How Small and Grassroots Nonprofits Can Tackle Evaluation and Talk Results**
Authors:	Marcia Festen and Marianne Philbin
Publisher/year:	Jossey-Bass (an imprint of Wiley), 2007
Extent/type:	136 pages, paperback
Price:	A$33.95/NZ$38.99 from Wiley Australia which offers a 15% discount to AES members, phone 1800 777 474 (within Australia), 0800 448 200 (from NZ only), +61 7 3354 8444 (from overseas), email <custservice@johnwiley.com.au>, website <www.johnwiley.com.au>
ISBN:	978-0-7879-7906-5

If you are an experienced and knowledgeable evaluator then this book is not for you! This is a lightweight (both literally and metaphorically) book for absolute novices without any evaluation background and who are about to launch into the world of evaluation within not-for-profit organisations. In fact, this slim volume forms part of a series of practically based books produced especially for this sector (covering, for example, topics such as fundraising and organisational development). The authors of this particular volume are consultants with considerable experience working with non-profit groups and have been joint authors before on how effective non-profit organisations work.

The basic goal of this text is said to be to 'help demystify evaluation and to offer practical strategies that enable more confident decision-making, sound program and organizational planning and increased accountability and credibility' (p. xiii). In addition, what is presented aims at the type of evaluation that (a) doesn't cost much (b) can be managed in-house and (c) doesn't require full-time, trained evaluation staff. In other words it is targeted at internal staff, board members and volunteers who have had no prior evaluation training or experience but see a need to introduce evaluation into the management process.

By teaching the basics of evaluation to such people, there is an underlying expectation presented by the authors that, gradually, an evaluation culture should become embedded within associated organisations, thereby enabling them to make a clear case for continued support/funding.

More specifically, the first chapter deals with the context in which current not-for-profit organisations work and why evaluation activity is important. While this section is interesting, it only considers an American context and so some material may seem irrelevant for Australasian readers. However, there are some useful sections on common fears and misconceptions for novice evaluators, and how to address these, as well as an indication of some poor reasons for considering evaluation.

The rest of the book then walks the reader through the evaluation process including matters to do with design, what to measure or assess, how to deal with the results, and how to put findings to use. These aspects are condensed into just five short chapters entitled 'Planning', 'Asking', 'Tracking', 'Learning' and 'Using'. Within these, there is plenty of step-by-step advice on how to work with staff, develop a plan, gather data and use results to best possible effect. The content is not dense with only 56 of the 120 or so pages being prose while the remainder is covered with supporting practical illustrations of forms, checklists, meeting agenda, case studies, and lists. This means there is variety in presentation and these elements are also presented in boxes with various shading, font sizes, etc.

In addition, there is subtle, continual, practical advice to support the underlying ideas. These are expressed in a way that

is simple and free from jargon. On the other hand, because the style is simplistic, there are some aspects of evaluation that are veiled or simply omitted. For example, even though there is talk of a 'theory of change' the term program logic is never used (except in an appendix), and while process and outcomes evaluation are discussed there is no mention of other types of evaluation, such as empowerment, goal-free or responsive evaluation. Then, in the section on data collection, methods suggested are merely the basic ones such as 'counting, observing, interviews, surveys and document analysis' and there is no mention of how to tackle analysis.

The end of the book is particularly disappointing. Problems occur because:

- there is no conclusion and so the reader is left dissatisfied with lack of closure
- the reference list is limited and there are some important basic evaluation books missing.

Also, the appendices listed as 'resources' are severely lacking, for example:

- there is a list of commonly used terms and definitions but there are only seven given (activities, benchmarks, data, impacts, outcomes, outputs, results)
- one appendix is entitled 'What is a logic model?', but it is not referred to in the text and no example/diagram is provided
- certain types of evaluation are listed as an appendix but there is no mention of forms such as monitoring.

In conclusion, this book can only be recommended if you do not want to be given any depth of understanding of evaluation for it only skims the surface of the discipline. On the other hand, if you are new to the field and only want a skeletal outline of what evaluation is about in 50 pages without getting bogged down in any detail whatsoever, then this may fit the bill!

Reviewed by:

Rosalind Hurworth

Director of the Centre for Program Evaluation
The University of Melbourne
Email: <r.hurworth@unimelb.edu.au>

Title:	**Diagnosing Organizations: Methods, Models, and Processes**
Author:	Michael I Harrison
Publisher/year:	Sage, Thousand Oaks, California, 2005, third edition
Extent/type:	181 pages, paperback
Price:	A$59.95/NZ$73 from Footprint Books which offers a 15% discount to AES members, phone +61 2 9997 3973, email <info@footprint.com.au>, website <www.footprint.com.au>
ISBN:	0-7619-2572-4

The third edition of Michael Harrison's *Diagnosing Organizations: Methods, Models, and Processes* represents a change in his approach to working with organisations to identify concerns and initiate change. In this edition he suggests that a more effective approach to diagnosing organisational effectiveness needs to be based on a more direct engagement with the client to identify the areas of direct concern. He also develops the role of change management as an integrated element of his approach in identifying how strategies for change can be best implemented within the operational context of the organisation.

Harrison suggests that there is a distinction between his diagnostic approach and that of evaluation practitioners where he sees their work as more focused and sustained investigations concentrated on specific program effects or program efficiency. In his book, Harrison suggests that his diagnostic approach seeks to gain a broad understanding of the organisations 'condition' by investigating a broad range of organisational parameters within a relatively short time frame. What is also emphasised is the pragmatic and responsive approach taken to meeting the client's needs and the provision of results that will be of immediate usefulness. With such comments if would appear that Harrison's approach may be more in line with the practices of many more evaluators than he thinks.

Through the use of the open system model, Harrison suggests the complexity of organisational structures where culture (norms,

values, beliefs), relationships between individuals and workgroups, technology and the environment provide a framework for understanding and investigating organisations. What is also stressed is the need to be aware that organisations operate within a 'political arena' based on the power and influence wielded by individuals and groups. While indicating that political actions are part of the way that organisations get the job done, Harrison also reminds us that when as consultants (or evaluators) we are asked to engage with an organisation it is the 'politics' of the organisation that will in some way shape any analysis and influence the way the findings are received and utilised. This may explain in part some of the focus of the final chapter that examines some of the ethical dilemmas of undertaking diagnostic work in organisations, including who benefits and who may be harmed by the process, and how the consultant addresses the issues of professional standards (and personal integrity), which can be maintained when these clash with the interests of the client.

Harrison indicates that the open model also suggest where the key for investigation should be in understanding organisations. In his view any analysis needs to be focused on individual and group behaviours and the organisational forces that shape these behaviours. To assist in this process Chapters 3–5 provide models for understanding individual and group behaviour, diagnosing congruence between system components and assessing organisational politics. These chapters are supported by quite detailed suggestions on techniques for gathering data. These are supplemented by the appendices that provide interview schedules, observation guides, and suggestions for a number of standardised instruments for use in diagnosis and assessment.

In what for me is a strength of his approach, Harrison argues that an understanding of the open systems framework and the broader cultural and political environment of an organisation can provide for greater utilisation (and impact) of the findings through assessing readiness and capacity to effectively implement.

While clearly based within the organisational and management theory paradigm, and with a focus on the external consultant as the key agent of investigation and change, the book provides a useful resource for understanding organisations and the techniques and tools that can be used for this purpose.

Reviewed by:

Graeme Harvey

Department of Education and Early Childhood Development
Victoria
Email: <harvey.graeme.lg@edumail.vic.gov.au>

Title:	**Organizational Ethnography**
Author:	Daniel Neyland
Publisher/year:	Sage, Los Angeles, 2008
Extent/type:	188 pages, paperback
Price:	A$73/NZ$88 from Footprint Books which offers a 15% discount to AES members, phone +61 2 9997 3973, email <info@footprint.com.au>, website <www.footprint.com.au>
ISBN:	978-1-4129-2343-9

This book promotes the use of ethnography in the study of organisations and takes the reader through the process of conducting ethnography in such settings.

The author provides a background to the development of ethnography in general, with its roots in anthropology, and in organisations in particular. Organisations relate to those settings usually studied by management research, although exemplars throughout the books draw on a broader range of studies. Neyland gives no indication of when ethnography may be the method of choice, leaving the decision to the reader on the basis of the discussion proffered in the book.

Methods and methodologies are considered through what Neyland terms 'sensibilities'. These are factors that need to be addressed before and during work with an organisation. There are 10 of these, with a chapter dedicated to each: ethnographic strategy, questions of knowledge, locations and access, field relations, ethnographic time, observing and participating, supplementing, writing, ethics, and exits. An

introduction and a conclusion make up the other chapters in this well-structured book.

- **Ethnographic strategy** refers to the need for the researcher to think about how (and sometimes why) they will engage with an organisation and go about gaining the insight they seek. It is pointed out that a prescriptive plan of action is not possible in ethnography and the researcher needs a fluid approach that may be subject to a series of changes. Many evaluators comfortable with a robust evaluation framework may find this approach problematical.
- **Questions of knowledge** provide a somewhat complex discussion, essentially addressing how information (knowledge) is to be garnered. Three principal approaches are put forward. In simplified terms, these are: realist, such as observation and access to informants; narrative, using a key informant to tease out and explain what is occurring; and reflexive, the use of participants in data collection, organisation and analysis.
- **Locations and access** involve decisions about which locations will be included in the study and how the researcher will get into, or engage with, the organisation.
- **Field relations** cover issues of how the researcher will interact within the organisation. There is a need to gain trust and establish close involvement, while recognising there are problems with being too close.
- **Ethnographic time** implies the need for immersion in the field for a long period and to proceed slowly. Neyland discusses the very long periods traditionally taken to undertake and report on ethnographic studies and contrasts this with the increasing demand for fast processes and measurable deliverables in management research. He points to the need for negotiation skills to ensure access for extended periods and an emphasis on the value of 'ethnographic time'. This would seem to suggest that the use of ethnography in organisational studies may be largely in the academic realm, a point that Neyland rejects. He defends the length of time and says that although this leads to suggestions that ethnography is expensive, this is not necessarily so. He notes that '(useful) ethnographic findings can often be provided at less expense than management consultancy' (p. 98) without saying how this might occur.
- **Observing and participating.** The ethnographer needs good observational skills and to be thorough in what is to be observed. Neyland notes that the longer the researcher participates in an organisation, the more familiar things become, increasing the potential for taking things for granted. The importance of field notes is highlighted, whether these are written during participation and observation or soon after.
- **Supplementing** suggests a number of strategies for gaining additional information, such as field interviews, cameras, and computers for virtual ethnography, gaining input through Web-based systems.
- **Writing** is discussed as occurring in two forms: scholarly pursuit or writing for the organisation.
- **Ethics** covers issues familiar to evaluators and includes some discussion on covert research, which is noted to breach a number of ethical principles.
- **Exits** need to be planned from an early stage and may be on the basis of the design of the research; for example, research may be of a phenomenon that is time limited, such as a merger or change of management. A study may cover a particular time frame within an organisation, for example a six-month or 12-month period. Organisational constraints may determine the time available or the evaluator may have particular time constraints.

Exemplars presented through the book are drawn on in the discussion of the various sensibilities. These are interesting and illustrative of the points made. In some instances, drawing parallels with management research takes some application of the mind. This does stimulate some thought about fresh ways of looking at how an organisation works.

Neyland makes the distinction between study *of* the organisation and study *for* the organisation. The former relates to setting research questions and solving them, while the latter is the application of research to be useful to the organisation. This distinction is visited a number of times in the book and finally, Neyland adds a third option, study *with* the organisation whereby members of the organisation undertake research that is coordinated by the ethnographer. While there is a discussion about utility, the book provides less guidance on moving to recommendations to assist an organisation to improve their processes and outcomes.

This book is 'designed as a guide to becoming an organisational ethnographer' (p. 8). It is set out in a clear, structured format and takes the reader through a logical sequence of development. As the author implies, it is most likely to appeal to someone who has developed an interest in and some understanding of ethnography.

Reviewed by:

Judith Woodland

Evolving Ways
PO Box 262
Ringwood, Victoria 3134
Email: <judith.woodland@evolvingways.com.au>

Title:	**How to Conduct Behavioural Research over the Internet: A Beginner's Guide to HTML and CGI/Perl**
Author:	R Chris Farley
Publisher/year:	The Guilford Press, New York, 2004
Extent/type:	299 pages, paperback
Price:	Price: A$49.95/NZ$61 from Footprint Books which offers a 15% discount to AES members, phone +61 2 9997 3973, email <info@footprint.com.au>, website <www.footprint.com.au>
ISBN:	1-57230-997-0

Overview

How to Conduct Behavioural Research Over the Internet provides a step-by-step guide to creating web-based surveys, as the basis for conducting behavioural research. A website <http://www.web-research-design.net> developed in conjunction with the book, contains all the CGI/Perl and HTML files used in the book that can be downloaded from this site. The website is an excellent companion to use alongside the book and there is a brief online quiz for each chapter, as well as a discussion forum, frequently asked questions, links to other relevant sites and an overview of the book including a synopsis of individual chapters.

Summary of chapters

Chapter 1, 'Introduction' outlines the aims of the book and what the reader will be able to accomplish by reading the book and completing the suggested activities. It provides a basic introduction to the Internet and definitions of key terms and concepts.

Chapter 2, 'Getting Started: A Step-by-Step Guide to Using a Web Server' outlines two ways to obtain access to a web server, either using a professional Web-hosting service or by building a Web server (and explaining how to do this).

Chapter 3, 'HTML Basics: How to Make a Web Page from Scratch' steps the reader through the creation of a web page in hypertext markup language (HTML).

Chapter 4, 'HTML Forms: Collecting Research Data from Participants via the Internet' discusses the use of HTML forms and how these can be used to obtain input/data from Internet users. It provides an overview of common Web-based input forms such as text boxes, radio buttons, checkboxes and pull-down menus.

Chapter 5, 'An Introduction to CGI Scripting: Using Perl to Automatically Save Response Data to a File' provides an introduction to CGI programming in Perl, the processing and storage of participant data, and the exporting of data to statistical packages. (CGI standards for common gateway interface, a standard for exchanging information between the user computer and the server.)

Chapter 6, 'Providing Customized Feedback to Research Participants' highlights the utility, of using Web-based research methods to provide feedback to participants. It outlines how to create CGI script to analyse and summarise individual participant responses and provide feedback to participants based solely on their own responses or in relation to the response of others.

Chapter 7, 'Randomizing the Order of Stimuli', explains how to use CGI script to randomise the order in which questions/items are presented while still maintaining the ability to process or store data in an organised manner.

Chapter 8, 'Random Assignment of Participants to Conditions', outlines how to randomly assign people to conditions by building on the randomisation techniques discussed in the previous chapter.

Chapter 9, 'Using Multiple Web Pages in Research: Carrying Responses Forward from One Page to the Next' discusses how to transfer/carry forward data collected on one web page to the next web page, through the use of 'hidden' HTML tags. This allows for questions, stimuli or sets of stimuli to be presented on separate web pages, while retaining the link to data collected on previous pages.

Chapter 10, 'Using Conditional Branching Structures: An Example of "Skip Patterns" in Survey Research', explains how to create conditional branching structures, using CGI script, where questions are presented to participants who meet a condition or set of conditions (e.g. if they are female, if they have children, if they have income over a certain level). Typically referred to as skips or skip patterns, these can be automated through the use of CGI script.

Chapter 11, 'Advanced Feedback: Summarizing Data with Bar Graphs and Two-Dimensional Plots' outlines how to summarise and present research results graphically using a combination of simple, preset images formatted in HTML.

Chapter 12, 'Tracking Participants over Multiple Sessions: PINs, Passwords, and Menus' explains how to store data from different surveys or data collections sessions for later use (e.g. comparisons at different points-in-time) by using personal identification numbers (PINs) and passwords to track respondent data over time.

Chapter 13, 'Measuring Response Times' discusses the contexts in which response times can and cannot be measured effectively over the Internet and provides examples to illustrate ways to assess response times online.

Chapter 14, 'Additional Applications of Perl: Discussion Forums and Scored Tests', discusses the application and use of CGI techniques for use in teaching contexts. It outlines

how to create online discussion forums and online multiple-choice tests that the server automatically grades and track completions and test scores.

Chapter 15, 'Wrapping it Up' concludes with a discussion of a range of miscellaneous but relevant Internet research topics including ethics, server maintenance, security, sampling issues, participant drop out, data quality control, and Web design.

Reviewed by:

Nan Wehipeihana

Research Evaluation
Consultancy Limited
PO Box 51313
Tawa, Wellington 5249
Email: nanw@clear.net.nz

Concluding summary

This text is a user-friendly, step-by-step guide to creating Web-based surveys and a range of excellent examples and resources are provided in the book as well as on the website.

At the time of being asked to review this book, I had just completed, with two other colleagues, a review of three online survey software packages. We were looking for an online software option that was cost-effective, was relatively straightforward in terms of setting up a survey instrument with an extensive and flexible range of question formats/response forms and most importantly, had excellent 'back-room' capability in terms of its utility and robustness for analysis and reporting.

Despite the clarity of this text I would not recommend that my colleagues build an online survey system to conduct research on the Internet, as outlined in this book. From the perspective of an evaluation practitioner, there are simpler and less time-consuming solutions, given the range of free or fee-based online survey products and services available.

The real value for practitioners lies not so much in the 'how-to' knowledge the book conveys, but the insight the book provides into understanding the issues that underpin undertaking online surveys (compared to other research methods). The book highlights the relative strengths and limitations of online survey and research methods and the factors to be aware of when designing or undertaking Internet-based research.

The clarity of writing, the format and layout of the chapters and the step-by-step instructional nature also make this an excellent reference text for students, or for use in teaching settings.

Title:	**Designing and Constructing Instruments for Social Research and Evaluation**
Authors:	David Colton and Robert W Covert
Publisher/year:	Jossey-Bass, San Francisco (an imprint of Wiley), 2007
Extent/type:	394 pages, paperback
Price:	A$72.95/NZ$82.99 from Wiley Australia which offers a 15% discount to AES members, phone 1800 777 474 (within Australia), 0800 448 200 (from NZ only), +61 7 3354 8444 (from overseas), email <custservice@johnwiley.com.au>, website <www.johnwiley.com.au>
ISBN:	978-0-7879-8784-8

I welcome this comprehensive resource on instrumentation for social research and evaluation practice. The authors develop a wide range of useful scaling and instrumentation techniques that are highly relevant for evaluation methodology. Their portfolio is more than sufficient for postgraduate courses on program evaluation.

According to the authors they assume and assert that:

> The process of constructing an instrument is both a creative and a technical venture. It involves not only being very familiar with the content or substance of the topic of interest but also developing good questions or items and presenting them in a format accessible to the people who will have to complete the instrument. Consequently, this book is designed to help you create an instrument that will obtain the information you seek. (Colton & Covert, 2007, p. xii)

Indeed they have followed this approach faithfully, for instance in the layout of the book with chapters on each stage of the instrument design, construction and use. However, with this approach lie both the book's strengths and weaknesses. It is avowedly both practical and non-technical, aimed at the reader with no technical background, minimising some aspects of the underlying theory, and even limiting the referral to where the theory might be elaborated. For example, it covers the essential issues of levels of measurement (Chapter 3), validity and reliability of instrumentation (Chapter 4), but it only gives a very short footnote on 'latent variable' with one reference, and nothing on Rasch Unidimensional Measurement Models, which some evaluation practitioners may wish to know about (see Bailey, 2001).

To give a bit more information about this excellent book I refer to seven criteria I have espoused in other reviews (Sharp, 1991). I propose that any publication of this kind (which purports to be a practical evaluation *handbook*) should meet seven criteria:

1 **Be comprehensive and eclectic.** This book does live up to this criterion, for example it deals with the wide variety of scaling techniques (including my favourite, Goal Attainment Scaling); and it deals with both quantitative and qualitative techniques; however, it is somewhat limited on the latest software (Chapter 14), indeed there is little on qualitative (content) analysis software (e.g. NUD*IST, see Richards & Richards, 1991).

2 **Offer a heuristic system.** Indeed there are many useful guidelines (e.g. Chapter 8 on writing selection items; Chapter 11 on constructing multi-item scales) with clear examples that illustrate these principles and also *what not to do* (as anyone with experience will agree it is important to point out bad examples of survey questionnaires to one's students and staff and to learn from these mistakes), as well as some helpful references to expertise, centres and systems.

3 **Provide a valid framework.** The authors do offer a research-based approach to instrumentation for data gathering and reporting in evaluation.

4 **Provide case examples from a variety of fields.** There is a rich store of examples from a diversity of fields, albeit with a bias towards the social sciences (e.g. consumer behaviour in market research, medical records administration audit, and employee satisfaction) and some evaluation examples (e.g. research quality evaluation, training needs, and training workshop evaluation).

5 **Be user-friendly.** This is not quite the student-friendly book I'd hoped for, although it is generally user-friendly in its systematic progression through the instrumentation process, and it is very easy to read and explains the issues and techniques well; thus, it is more of a beginning (to somewhat experienced) practitioner's handbook than a class textbook; it does have useful contents, and indexes; however, although the authors provide a list of key terms at the end of each chapter, a glossary and homework exercises would have been better, making it more convenient for use by the teacher and the student; given this approach it may be quite acceptable that the references are minimal and focused on the practitioner's needs, rather than elaborating on the history and theory of the tools and techniques.

6 **Be technically competent.** There is no doubt that Colton and Covert have demonstrated their command of the research and technical aspects of this field in this book, which is based on many years of relevant experience, to deliver very sound methods and approaches.

7 **Provide an adequate explanation of data collection and analysis.** Undoubtedly, the authors have conveyed the relevant information on data collection instruments for the uninitiated. They do well in covering not only the basics but also the '*creativity*' or *thinking* involved in the design and use of instruments for data collection, analysis and reporting for social research and evaluation.

Overall strengths and weaknesses

The main strength of this publication is its accessibility as a useful handbook with clear reliable guidelines for designing tools for data collection for evaluation. Its weakness is that it may be too focused on the designing of the instruments when the student of evaluation may be looking for more (especially on computerised and Internet-enabled techniques) in relation to the qualitative as well as quantitative methodologies of analysis of the evaluation data. These issues of analysis are introduced but are not covered as comprehensively (nor in as much detail) as the basic focus on instrumentation.

Overall, I applaud and recommend this book for the evaluation practitioner and I have used it as a handy supplementary text for my postgraduate courses and workshops.

References

Bailey, S 2001, 'Measuring customer satisfaction: comparing traditional and latent trait approaches using the Auditor-General's client survey', *Evaluation Journal of Australasia*, new series, vol. 1, no. 1, pp. 8–17.

Richards, T & Richards, L 1991, 'The NUDIST qualitative data analysis system', *Qualitative Sociology*, vol. 14, no. 4, pp 1 ff (see also <http://www.qsrinternational.com//default.aspx>).

Sharp, CA 1991, 'The Program Evaluation Kit, 2nd edition', book review, *Evaluation Journal of Australasia*, 1991, vol. 3, no. 1, pp. 46–50.

Reviewed by:

Colin Sharp

Fellow of AES;
Managing Director
P.E.R.S.O.N.A.L. (Research & Evaluation) Consultancy Pty Ltd
PO Box 378
Stirling, South Australia 5152
Email: <PERSONAL.research@bigpond.com>

NOTES TO CONTRIBUTORS

All AES members and other people involved in the practice, study or teaching of evaluation are invited to submit articles, reports, reviews or news items for publication. Before submitting articles, contributors are advised to consult the full guidelines set out on the AES website (**www.aes.asn.au**).

Decisions on publication are made on the basis of assessments by the editors, taking account of advice from authorities in relevant fields. Where requested by the author, articles will be submitted for formal peer review, and such refereed articles will be identified accordingly in the journal. However, the editors reserve the right to publish a paper as non-refereed or refereed as appropriate, in consultation with the relevant referees.

There is no predetermined 'house style' for *EJA*. Authors are encouraged to give expression to their own individual voice, provided it meets generally accepted standards of clear, accurate and literate English usage. The principles set out in the *Style Manual for Authors, Editors and Printers* (sixth edition) should be used as a general guide to usage. Citations should follow the Harvard (i.e. author–date) style.

Authors are asked to write their articles in a way that foregrounds the most significant and interesting features of their content, avoiding unfamiliar technical language, extensive data tables and mathematical notation, except to the extent that these are necessary to convey the argument. Articles should be arranged in a way that facilitates comprehension by non-specialists, e.g. by using text boxes for background information and aspects of detail that are not central to the exposition. References should be kept to the minimum required for understanding the text or following up sources specifically quoted or discussed in the article.

The editors reserve the right to correct obvious errors in spelling, vocabulary and grammar without reference to the author. Any more extensive corrections or modifications which the editors and/or peer reviewers consider desirable will be discussed and explicitly cleared with the author before being incorporated in the published text.

Individual contributions should not normally exceed a published length of 10 pages, including all tables, graphics, notes and references. The editors reserve the right to edit longer articles down to this length if space considerations require. Articles shorter than 3000 words will not generally be regarded as suitable for peer review.

Copyright in articles submitted for publication in *EJA* resides with the authors, unless otherwise specified. Copyright in the form of the articles as presented in *EJA* resides with the AES.

Articles should be submitted as attachments to an email, preferably in Word (with the extension .doc) or as rich text format (RTF) to:

Associate Professor
Rosalind Hurworth
Director
Centre for Program Evaluation
Faculty of Education
University of Melbourne VIC 3010
Tel: +61 3 8344 8624
Fax: +61 3 8344 8490
r.hurworth@unimelb.edu.au

or

Dr Delwyn Goodrick
Evaluation Consultant
Mobile: 0438 388 204
delwyngoodrick@gmail.com

Items for publication may also be sent to:

AES Office
PO Box 5223
Lyneham, ACT 2602
Fax: +61 2 6262 9095
aes@aes.asn.au
marked 'Attention: EJA Editors'

Full guidelines for contributors are published on the AES website (**www.aes.asn.au**) in the 'Publications' section.

COPY DEADLINES

Vol. 9, no. 1

24 April 2009

Vol. 9, no. 2

21 August 2009

Intending contributors who do not expect to meet these dates are asked to contact the Editors as soon as possible, but prior to the nominated deadline.

AES BOARD

Up-to-date details at www.aes.asn.au/about