METHODOLOGICAL ISSUES IN WORKING WITH POLICY ADVISERS AND PRACTITIONERS

by

Gloria Laycock
University College, London

Abstract: This chapter considers what policy advisers and practitioners arguably need from researchers, and what they can reasonably expect. It goes on to discuss the implications of these needs and expectations for the methodologies used in research. But we begin with some definitions — what we mean by policy adviser, practitioner and researcher in this context, and why they need definition.

DEFINING TERMS

It may seem odd to begin a chapter with the definition of such basic terms as policy adviser, practitioner and researcher, but there are some specific characteristics of these individuals that are central to their expectations and their capabilities and it is, therefore, worth spelling these out. This is particularly true in the context of crime prevention, where the assumptions and beliefs of the general public form the backdrop against which policy makers and practitioners operate (Tilley and Laycock, 2000). They would not be performing their jobs properly if they were to ignore the public view. The same is far less true of the researcher, who, in theory at least, should be trying to acquire knowledge independently of the public's views.

So let us begin with the "policy adviser." In the United Kingdom we are talking about senior civil servants, in the United States of America, political advisers at federal, state or even local levels whose responsibility it is to advise the governors and legislators about crime

and its control. They are one step removed from politicians but are nevertheless aware of the political pressures operating at that level and, therefore, well advised to attend to the expectations and beliefs of the public. In making sensible recommendations about crime prevention many will be conscious of the long held public view that the police, acting through the criminal justice system, can and should be expected to reduce crime. And some will share that view. It is an approach that generally leads to an expansion of the enforcement model — it fuels cries for more police officers on the streets, and for tougher and longer sentences. It does not point, naturally, to the notion that environmental measures play a major part in generating criminal or disorderly behavior. So any advice reinforcing, for example, the role of opportunity in crime control (Felson and Clarke, 1998), has to overcome these beliefs as a first step. This message has the added disadvantage that it is not cost-neutral to the recipients. The members of the public, and a wide range of agencies in commerce and industry, have to take some responsibility for crime control — they actually have to *do* something rather than sit back and expect the police and others to protect them (Laycock, 1996). This is not an easy message politically. It involves, at least partially, handing crime control back to the people from the State, with the accompanying responsibility (Christie, 1977).

The practitioner can be characterized as an individual who is expected to deliver the policies "on the streets." In the case of crime prevention, this is often the police, although, depending on the scope of the crime prevention effort, it might also include the other statutory or voluntary partners with whom they are now beginning to work (Crime and Disorder Act, U.K., 1998). They, quite literally, come face to face with the beliefs and expectations of the public in a way in which the more remote policy advisers do not. For the purposes of discussion in this chapter, we are restricting the definition to statutory police agencies — those whom the general public see as central to the delivery of safe and crime-free communities, although the extent to which these expectations can be met by the police alone has been challenged in official U.K. publications for about two decades and the idea that the police need to work in partnership is also central to the current "community policing" approach in the U.S.

Finally, the researchers — how are they to be described? There is a sense in which a wide range of activities can legitimately be called research. The work of a detective, for example, might be so construed. He or she will go through a systematic process of gathering evidence and apply various rules of logic or systematic review to try to achieve a satisfactory solution to a case. And a wide range of disciplines might claim to carry out research in the normal course of

their work — the journalist, historian, or scientist — all use systematic processes, with claims to objectivity and the application of scientific method. In this chapter we are taking a fairly narrow definition. The researchers as discussed here will draw on the social sciences, will use some sort of data analysis, quantitative or qualitative, in the course of their investigations, and, importantly, will formulate and test hypotheses about crime and its prevention. They will be familiar with the research literature, or know how to access it. They will also draw on the various theories of human behavior from psychology, sociology or the behavioral sciences more generally in their efforts to describe, understand or test the crime prevention programs with which they may be associated. They will not, therefore, only be concerned with producing descriptive statistics, or even the evaluation of programs, unless there is a clear statement of the process or mechanism through which those programs might be operating. To do otherwise risks producing barren reports which tell the knowledgeable adviser or experienced practitioner what they already knew. The challenge to the research community is to tell the world something new and take the field forward, rather than looking back, as so often happens (Christie, 1997).

WHAT DO POLICY ADVISERS AND PRACTITIONERS NEED FROM RESEARCH?

Bearing in mind where these people are coming from, as they say, what might their needs be? Some suggestions are listed below under the separate headings of policy makers and practitioners, and the section concludes with a brief discussion of some of the constraints on the researcher in being able to meet the demands.

Policy Advisers

The world in which the policy adviser works is a fast changing and demanding one. This is a common feature of the scene in both the U.S.A. and the U.K., although there are some major differences between these two jurisdictions — in the U.S.A. senior policy advisers are often appointed by politicians and may, for example, have an implicit remit to support the politicians' re-election or at least to toe the party line. The advisers are, therefore, not only sensitive to whether the messages from research fit with the public's expectations, but they are also expected to address the extent to which research results are compatible with the political complexion of their patrons. In the U.K. the position is somewhat different. Although there are politically appointed advisers in all government departments, and they

have grown in number over recent years, they are still vastly out-numbered by the permanent senior civil service whose advice is expected to be politically impartial and factually based.

In both the U.S.A. and the U.K. recent administrations have begun to look for "what works" as a means of improving efficiency and reducing costs. This interest has been something of a two-edged sword. On the positive side, it has led to substantially increased demand for research, with the associated funding and implication that policy might finally be influenced by it (U.K. Home Office, 1999; U.S. National Institute of Justice, 1997). More unhelpfully, it has also exposed the amount of money that has already been spent on research, and the relatively little by way of firm evidence on effectiveness that it has produced (Sherman et al., 1997; Goldblatt and Lewis, 1998).

In both jurisdictions, however, it is probably fair to say that the advisers are looking for "good news," i.e., the policies that were effective — in the crime prevention world, this means that crime went down, the public was less fearful, everyone was happy. So the first somewhat cynical requirement from research is that it produces what the politicians want to hear. And it needs to do so quickly. There is an impatience about the process of government that leaves research at the starting block. While the average university professor is beginning to contemplate the enormity of the research exercise facing him or her, the politician is expecting an answer, and an unequivocal one at that. Conclusions that "more research is needed" receive a very bad press.

Despite their wish for good news, and the rush to announce new initiatives, the policy adviser in this new and developing outcome-focused world will increasingly need to have confidence in the conclusions of the research exercise, particularly if large sums of public money are at risk of commitment to new programs. In the "old days" it mattered less if decisions were made on the basis of political expediency rather than efficacy, but now, with the public, the media and others watching the public sector equivalent of the "bottom line," confidence in research conclusions matters. And an additional dimension has been added with the interest in reducing the cost of central government activity — not only do policy advisers want to know what works, they want to know at what cost. Cost-effectiveness is increasingly being built into the evaluation plans for new studies (Dhiri and Brand, 1999; Colledge et al., 1999).

Next is the delicate matter of money. There is a tendency for policy advisers in central government to see themselves as the custodians of public money — and he who pays the piper expects to call the tune. In the U.S.A. the process is more directly dependent on the views of Congress, where research money is voted for specific purposes, with

at present a relatively small amount for discretionary use. The degree of flexibility within the discretionary allocation varies, but the head of the federal government's criminal justice research agency in the United States is a presidential appointee, and is expected to take the government's priorities, and view of the world, into account when setting the agenda.

In the U.K. the position is again rather different. Here the money voted by Parliament for centrally funded criminological research has generally been for the support of the criminal justice policy process. The policy advisers were, therefore, seen as the customers for the work (although ministerial approval was generally expected) and felt that they should have the major say in research expenditure (Cornish and Clarke, 1987). More recently the picture has changed somewhat, with the "legitimate user" being expanded from the policy advisers to include the practitioners, and with more weight being given to the experience of the government researchers and their views, in setting the research agenda. The relationship is now better characterized as a set of partnerships between policy, practice and research (Laycock, 2001). Whatever the subtleties of the developing relationship between researcher and policy advisers, the policy advisers feel they have a significant role to play on both sides of the Atlantic — and that they should not, therefore, be ignored in determining the research agenda. While this partnership approach dilutes the political influence over the research agenda, in reality it does not remove it. Political ideology at least maintains a "filtering" role (Doherty, 2000) over what research is done, how it is presented publicly, and what influence it may have over future policy direction.

Finally, and in some ways this ought not to need saying, policy advisers need to be able to understand the results of a research project. Research papers that are written in obscure technical language, permeable to the chosen few, and covering reams of paper, are less likely to influence policy than a concise, crisp few pages which summarize the important points and spell out the implications for policy. Some researchers are distinctly uncomfortable with this scenario, particularly the idea that they should perhaps go somewhat beyond their data in spelling out the policy implications of what they have done.

The requirements of the researcher from the policy adviser can therefore be summarized as:

- Good news;

- Confidence in the results;

- Costs included in evaluations;

- A feeling of involvement in the agenda setting process;

- Timely production of results;

- The identification of "what works";

- Good communication skills; and

- A willingness to take risks in making inferences from their data.

Practitioners

The world is changing rapidly for criminal justice practitioners. It is no longer good enough to point to a list of outputs from the field — more tickets issued, more arrests made, more neighbourhood/block watch groups or partnerships established. The pressure is on for the delivery of outcomes. There are a variety of ways in which this is manifesting itself. In the U.K., for example, the Crime and Disorder Act requires the police and local government to work together and produce a strategic plan for the reduction of crime and disorder. There is an increasing expectation that this will be related to hard targets for reduction. Indeed the Government has set itself a target of reducing vehicle crime by 30% over five years.

In parallel, still in the U.K., a new regime of "best value" has been introduced under which police forces are expected to examine their policies and procedures on a regular cycle and demonstrate that they are achieving value for money. This does not mean going for the cheapest option necessarily, but going for the most cost effective (Leigh et al., 1999).

In the U.S.A., one of the more tangible manifestations of this move toward an outcome focus is through the Compstat process first used in the New York City Police Department and now being replicated in a number of police agencies across the country. Compstat is no more than a management tool forcing the attention of precinct commanders to the crime and disorder statistics in their area. Typically there will be a map of the precinct with the recent crime and disorder incidents displayed. The commander will be expected to give an account of what he or she intends to do about the offending pattern. How will it be reduced?

In both the U.S.A. and the U.K., there is also a significant move toward problem-oriented policing (Goldstein, 1979, 1990) — the police are expected to solve problems rather than to react continually to them. Again there is an expectation that what might have been a recurring problem will be addressed in such a way as to reduce the

rate at which the incidents occur, or to eliminate them — a focus on outcome rather than output.

One collective consequence of all these developments is the increasing pressure on practitioners to implement cost-effective tactics to reduce crime. There is, therefore, a need for the practitioners to know what works, where, and at what cost. They are looking to the research community to provide some answers.

In this respect the picture has changed dramatically over the last five or so years. In the past, the researchers and practitioners could operate in independent universes for all practical purposes. The researchers could carry out their research *on* the police and report the results in their academic journals. The police could complain about the unfairness of the results and the frequent negative findings — nothing works — and ignore the outcome. This is clearly a caricature of the situation, but it is not that far off. If we calculate the amount of money spent on police-related research in the U.S.A. over just the past decade, and then look at how much it has impacted operational policing, then we might wonder where the money went. Although some commentators argue that the influence has been subtle but important, on, for example, police responsiveness and accountability (Bayley, 1998), the view taken in this chapter is that at best it provides poor value for money — if, that is, changing the way the police do business was the intention of the research, and it often was not.

Like the policy makers, the practitioner should be involved in the agenda setting process. Although they do not have resources to spend on research to the same significant extent, a great deal of research time can be saved if practitioners are involved in the process of agreeing to the research program from the beginning. Research hypotheses can be generated on the basis of their experience for example. In addition, because of their proximity to "the streets," they are often the first to become aware of a rising crime problem, which in the normal course of events might take months or even years to permeate the policy consciousness. An ability to pick up these emerging problems early would be one advantage of involving practitioners in determining the research program. Not only could the emerging problem be quantified and described, but a timely research-based response could nip it in the bud.

Many practitioners also appreciate some research support in tackling problems unique to them. They are on the receiving end of policy decisions, and are often ill advised on how, for example, the decisions might best be implemented and what others are doing to deliver the latest proposals in a cost-effective and timely manner. They need to know not only what works, but where and why. Neighborhood or block watch is a good example of an initiative that is

highly sensitive to the context within which it is introduced in what it can deliver (Laycock and Tilley, 1993). Practitioners need to be aware of these subtleties.

Perhaps most importantly, what practitioners need to understand is *why* an initiative may or may not have worked. Knowing this will help them, and their research colleagues, in deciding where and how to replicate the initiative (Tilley, 1993), and ultimately to "main-stream" it. At some point the research process needs to lead to the articulation of the "mechanism" (Pawson and Tilley, 1997a) through which the initiative is presumed to be effective. This amounts to saying what the hypothesis is that might lead to a reduction in crime. In the U.S. literature this is called the "logic model" or "program the-ory" (Weiss, 1998a).

Like policy advisers, practitioners also need to be able to under-stand the results of the research process. Results need to be pre-sented in plain English, with the practical implications of projects written clearly. Busy practitioners do not have the time, nor in many cases the interest to immerse themselves in the details of research. A different product is required which meets the needs of the practitio-ner audience.

Finally, and perhaps a relatively trivial point, practitioners appre-ciate (even if they do not "need") feedback when they have given their time, resources or effort to support a research exercise in their agency. Too often there are complaints that researchers come and go — collecting data, issuing questionnaires or generally making de-mands, without providing any feedback on the final results of the exercise.

To summarize the requirements of the practitioner, they need:

- To know what works, where and why;

- Help in replicating "what works" — understanding contexts and mechanisms;

- Help in generating testable hypotheses;

- Timely research;

- Involvement in setting the agenda;

- Reports and recommendations in plain English;

- To know of current good practice; and

- Feedback on the results of research in which they have par-ticipated.

None of the requirements outlined above are in any way illegiti-mate or unfair. They simply reflect the realities of life. If research is

to be of use to policy advisers and practitioners then it needs to be attentive to the needs of these communities.

Constraints on the Researcher

If these expectations of research are fair and reasonable, why have they not been met as extensively as they might? There are a number of possible reasons, largely stemming from the way in which research is organized and funded both at national level (where decisions about expenditure are made and where the focus is on project management) and within the university or consultancy sector (where the bulk of the research is carried out). There are also reasons related to the way in which researchers in social science have typically been trained. Perhaps the primary reason, however, is simply that government-funded research was no more clearly outcome-focused than the processes and procedures that were the subject of that research. And there was no requirement that it should be.

The bulk of the research expenditure resulted in published articles, academic books, briefing notes, conference papers — all output measures, not outcomes. Indeed, outcome measures for research are particularly difficult to develop (Weiss, 1998b), not least because few research exercises have been designed explicitly to influence practice. Researchers, like policy advisers and practitioners, have been working to their own agenda. In their world, they are rewarded for articles published in refereed journals, for the number of citations they amass, and for attracting research grants. They are concerned with tenure policies and, at a professional level, are rightly expected to attend to the quality and professionalism of their work. There are no brownie points for sticking the academic neck out and speculating beyond the collected data set about the implications for either policy or practice of the research exercise on which they may have been engaged. And some might say quite right too.

Researchers are also wary of being seen to be "too close" to either the policy adviser or the practitioner. On the policy side they may be vulnerable to pressure to come up with a good news story, which supports the party line. On the practitioner side they may be influenced by personalities, be caught up in the detail of the local initiative or, for other reasons, simply find themselves investing personally in the outcome of the study and produce bias in their results as a consequence.

All of these issues and concerns affect the way in which researchers carry out their tasks, including their choice of methodology. The remainder of this chapter will discuss some of the methodological implications of what might be called a "new agenda" (Kennedy, 1999)

for researchers. This new agenda will need to meet researchers' needs, but also those of their "partners." It will not consider some of the other issues, which are equally relevant to the implementation of research results, such as the commissioning process and the presentation and dissemination of results.

IMPLICATIONS FOR METHODOLOGY

By way of a reminder, we are looking at methodologies relevant to determining what works (for the policy advisers) as well as where and why (a major concern of the practitioner). These methodologies need to be acceptable ethically and politically — by which we mean that there is no sense in suggesting a randomized control trial on the deterrent effect of removing body parts of convicted offenders, no matter how effective or even cost-effective such a process might be felt to be.

Practitioners want to feel part of the research process rather than the remote subjects of it. They want to be included, not excluded, and they want their views and experience taken into account. They do not want to be told, at the end of a two-year evaluation, that there was an implementation failure from day one — but nobody told them. And the policy advisers do not want any surprising bad news — if things are going badly then the sooner they hear about it the better — and in any event they do not want to read about the bad news (or even the good) in their morning paper.

Policy advisers, in particular, are concerned to know as quickly as possible whether or not an intervention is proving effective. This is because they are often under pressure to "mainstream" the initiative almost as soon as the pilot site has been announced. This expectation needs to be "managed." Interim results might be helpful, for example, or at least some early lessons on implementation.

Against this background, the remainder of this chapter first considers "experimental methods" (randomized controlled trials and quasi-experimental designs), which have been described as setting the gold standard for evaluation. It then goes on to consider some alternatives to experimental methods, which, it is argued, better meet the needs of the time.

Experimental Methods

Experimental methods have an impressive pedigree. The advancement of knowledge in the field of medicine, for example, owes much to designs of this type, where the question is the extent to which particular medical treatments may or may not "work." As such, these designs are obvious candidates for transfer into other

areas of social policy where similar "what works?" questions are being asked. The medical experience is often seen as setting the standard to which other social policy research exercises might aspire. There are, though, some major differences between the way in which medical issues can be addressed and the problems faced in other social policy areas. In medicine, the treatments themselves have, traditionally, been fairly readily defined in most cases — dosages can be prescribed, for instance, so as to make "implementation failure" unlikely. Also, although the picture is now becoming more complex, there is generally a relatively clear outcome — the patients' conditions improved or they did not (defined as mortality, morbidity, quality of life and/or patient satisfaction). In addition, the unit of analysis is typically an individual, and it is feasible to allocate individuals to treatment, control or placebo groups on a random basis. The sample sizes can be manipulated to reflect the expected effect, with large samples being used, for example, where treatment effects are expected to be marginal. It is also feasible to carry out more than one such randomized controlled trial (RCT) in any given research field. Furthermore, there is a reasonable degree of certainty that sufficient numbers of patients will follow the treatment regime to make the trials viable.

Furthermore, it is sensible to test the hypothesis that treatment X, a particular dosage of medication, will be to some degree effective across all patients. While there is assumed to be individual variation in the appropriateness of the treatment, the assumption is that it will be more effective than not on aggregate. In other words, there will be a relatively small number of patients for whom it may be marginally less effective, but they will be sufficiently small in number not to affect disproportionately the outcome when compared with the untreated control group. Where the delivery of the treatment is heavily dependent upon the skills of an individual — in surgery, or psychotherapy, for example — the experimental procedure may need to be more complex, and larger samples may be required, if the effects of the different individual styles are to be controlled for.

It is probably also fair to say that in many if not most of the early clinical trials, researchers were comfortable with the notion that the mechanism through which the treatment was expected to exert its effect was unknown, and it was not necessary that it should be known in order for the experiment to be carried out "successfully." So, for example, the efficacy of penicillin could be tested using randomized controlled trials without any real understanding of *how* it was delivering its effect. As knowledge develops in relation to such mechanisms, it becomes possible to test efficacy in a more sensitive manner: hypothesizing, for example, that the effects may be greater

for some sub-sets of the population than for others. And since samples can be relatively large, it is also possible to test for interaction effects in the course of a standard experimental trial.

Whilst RCTs might, to varying degrees, be an appropriate research model in the medical arena and in some other areas of public policy, there are some major differences in the crime prevention/policing fields, which raise serious doubts about their usefulness there. First, the unit of analysis is often not an individual; it may be a store, community, housing area or parking lot. Random allocation of a sufficiently large sample of stores, communities etc. to a treatment and control group is, generally, quite impractical.

Secondly, RCTs are "black box" experiments where the mechanism through which the effect is taking place is not necessarily known. So even if we were able to allocate randomly a large sample of communities to treatment and control groups, such allocation would not necessarily provide any information on *why* the initiative may have been successful. The outcome of the experiment is, therefore, along the lines that treatment X may have worked, but we do not know how or why. A conclusion of this kind is not what is required at this stage in the development of knowledge about policing and crime prevention.

Thirdly, RCTs assume that there is relatively little risk of implementation failure, or if it does happen it is in sufficiently small a proportion of the experimental population as to be irrelevant. In policing and crime prevention at the community level, implementation failure is a significant possibility, and major efforts have to be made to ensure against it (Laycock and Tilley, 1995).

Finally, in RCTs the experimental treatment has to be maintained throughout the experiment, as originally intended; learning from experience, or making ongoing improvements or adjustments is not permitted. While these assumptions and constraints are plausible in the medical field, with sufficiently large samples, they are not tenable in communities where there is far less control. Crime moves, evolves and develops with the opportunity structures offered by the immediate situation within which the potential offender finds him or herself, and which are beyond experimental control. For all these reasons, then, true experimental designs of the kind ideally deployed in the medical field are generally neither practical nor desirable in the real world of policing. There are far too many variables to even pretend that they can be controlled.

Quasi-experimental Designs

Quasi-experimental designs, employing experimental and control neighborhoods, and where there is no attempt at random allocation, are a rather different matter, although even here the reality necessarily falls somewhat short of the ideal. These designs require a control area, and possibly a displacement area, with which to compare crime rates in the experimental area where the innovation under investigation is taking place. The control area is meant to be comparable to the experimental area in social composition etc., but is expected to remain free of any innovation or other attempts to affect the dependent variable (the crime rate). The displacement area, usually adjacent to the experimental zone, is meant to test for the possibility that instead of reducing crime, the experimental effect was merely to relocate it (Barr and Pease, 1990). Crime figures over a reasonably long period are collected before and after the experimental intervention in all three areas.

There are some conceptual difficulties with such an approach. First, it is assumed that nothing other than the experimental treatment will affect the dependent variable in the experimental area, and nothing at all will happen in the control or displacement areas that might affect the dependent variable — or that if anything does affect the dependent variable it does so in all three areas (experimental, control and displacement) in equal measure. This is an almost impossible condition to guarantee in most real world experiments in communities where there are all sorts of changes and dynamics operating.

It also assumes that the crime rates in all three areas — experimental, control and displacement — are reasonably stable over time. Too much random fluctuation, or "noise" would make any empirical investigation questionable. In reality, the relatively small areas that are usually subject to experimental investigation are susceptible to random fluctuation, and are very difficult to control in the sense of ensuring that no other relevant activity is spontaneously occurring at the same time as the experimental investigation. Thus, it is not that unusual to find that crime has been reduced more in the so-called "control" area, for reasons that may well be beyond the experimenter to explain. And, as with true experimental designs, the experimental treatment has to be maintained in its original form throughout the exercise and not modified in the light of experience.

To make matters worse, the notion of "diffusion of benefits" has been developed in recent years (Clarke and Weisburd, 1994). Instead of crime moving from the experimental area to the adjacent "displacement" area, the benefits of the experimental intervention may

spread into the displacement area and reduce crime there too. It is possible to offer some plausible speculations as to why this may happen, but the fact that it is a possibility produces problems for quasi-experimental designs.

Finally, the kinds of interventions which practitioners are now claiming might be effective in reducing crime or in improving policing are often multi-faceted. They suggest that a package of interventions is required in order to turn around disadvantaged communities. Approaches of this kind are particularly difficult to evaluate no matter what the approach, since it is not possible to determine which, if any, of the experimental interventions produced the observed changes; but they are certainly not susceptible to evaluation using quasi-experimental designs.

The criticisms of experimental and quasi-experimental designs have not, so far, been directly related to the extent to which they can provide the kind of information that is required by policy advisers and practitioners. They fail on a number of criteria in this area also. These designs are expensive and time consuming, but they are also unrealistic in expecting the control and displacement areas to remain intervention-free, and in expecting practitioners to ignore the incoming feedback — which may suggest the need for changes to the project — as an intervention begins to take effect. Finally, by their very nature they are not able to explain why an initiative might be working or what the context sensitive dimensions of it might be. All these issues go some way toward explaining the relative rarity of such experimental projects in the criminal justice field, and their inappropriateness.

This is a fairly depressing picture for those aspiring to evidence-based policy and practice. The picture is not, however, as gloomy as it may appear. Over the past few years a number of alternative approaches have been developed in other disciplines which are transferable to policing, and new approaches have also been tried within the criminal justice field itself, as discussed below.

Alternative Approaches

There is a huge range of alternatives to experimental methods. It is beyond the scope of this paper to go into the detail of how each of these might be used to develop and evaluate programs and projects; there are, however, three techniques, which seem particularly well suited to the kinds of evaluations now required. They have certain features in common, and after a brief description of each approach, their common features are discussed.

Theory of Change Approach

This approach has a fairly long history; indeed it could be described as experimental psychology. In recent years it has taken on a new label and has been adopted beyond psychology as an approach to evaluation more widely. A complex discussion of "theory-driven evaluations" is provided by Chen (1990), which draws on earlier work by Chen and Rossi (1987) and which has much in common with the theory-of-change approach more simply described by Weiss (1972, 1998a) and later by Fulbright-Anderson et al. (1998).

The impetus for the present incarnation of this approach derives from the need to evaluate comprehensive community initiatives (CCIs), which the traditional experimental approaches are ill designed to do (Kubish et al., 1995). CCIs are difficult to evaluate because they are complex, they are trying to solve a number of possibly related problems at once; they are flexible and evolve as the problems themselves change, and there are no control groups with which they can be sensibly compared.

In order for these complex interventions to be evaluated, the program manager or sponsor has to be specific about how and why the program, in its various aspects, might work (Weiss, 1995). The evaluator, with help from the program staff, has to be able to articulate the theories, assumptions and sub-assumptions that might be built into the program. It is these that form the hypothesis or set of hypotheses, which the evaluation will go on to test. Some of these theories can be complicated and inter-related, making firm conclusions and generalization difficult. In some cases, the practitioner may not be clear about how or why the program is expected to work, and they may need help in working through the possible alternatives. They may just *know* it is the right approach and resent the analytic approach taken by evaluators in pressing them to work through the underlying theories. There may also be some disagreement among practitioners or between them and the program sponsors. Finally, there is a possibility that being explicit about just how something might have an effect may be politically problematic. For instance, the mechanism may be interpreted as divisive of the community, or may appear to favor one sub-group over another. Whilst such possibilities can be glossed over when the program assumptions are not brought out, it is much more difficult to do this when they are on the table for all, including the media, to see.

If, on the basis of a number of specific theories, a program appears to have worked, how far can we generalize to other communities in what might be quite different areas? The fact that the theory has been set out is helpful in answering this question, and the extent

to which it is capable of generalization depends upon the theory itself. If, for example, it relates to some very specific attribute of the particular community, or to the skills or inter-personal relationships of central players, then the scope for generalization may be less. If the initiative is to be shown to be more generally applicable then it will be on the basis of its replication in other areas. In order for replication to occur, the theory has to be clearly stated. Replications are notoriously susceptible to replicating the wrong thing (Tilley, 1993).

Let us take a specific example from the U.S. literature. Weiss (1998a) describes the program theory behind the proposal that higher teacher pay may increase student achievement. There are a number of possibilities why this may work, as set out in Figure 1.

To quote Weiss (1998a:57) directly:

> program theory, ...refers to the *mechanisms* that mediate between the delivery (and receipt) of the program and the emergence of the outcomes of interest. The operative mechanism of change isn't the program activities per se but the response that the activities generate (original emphasis).
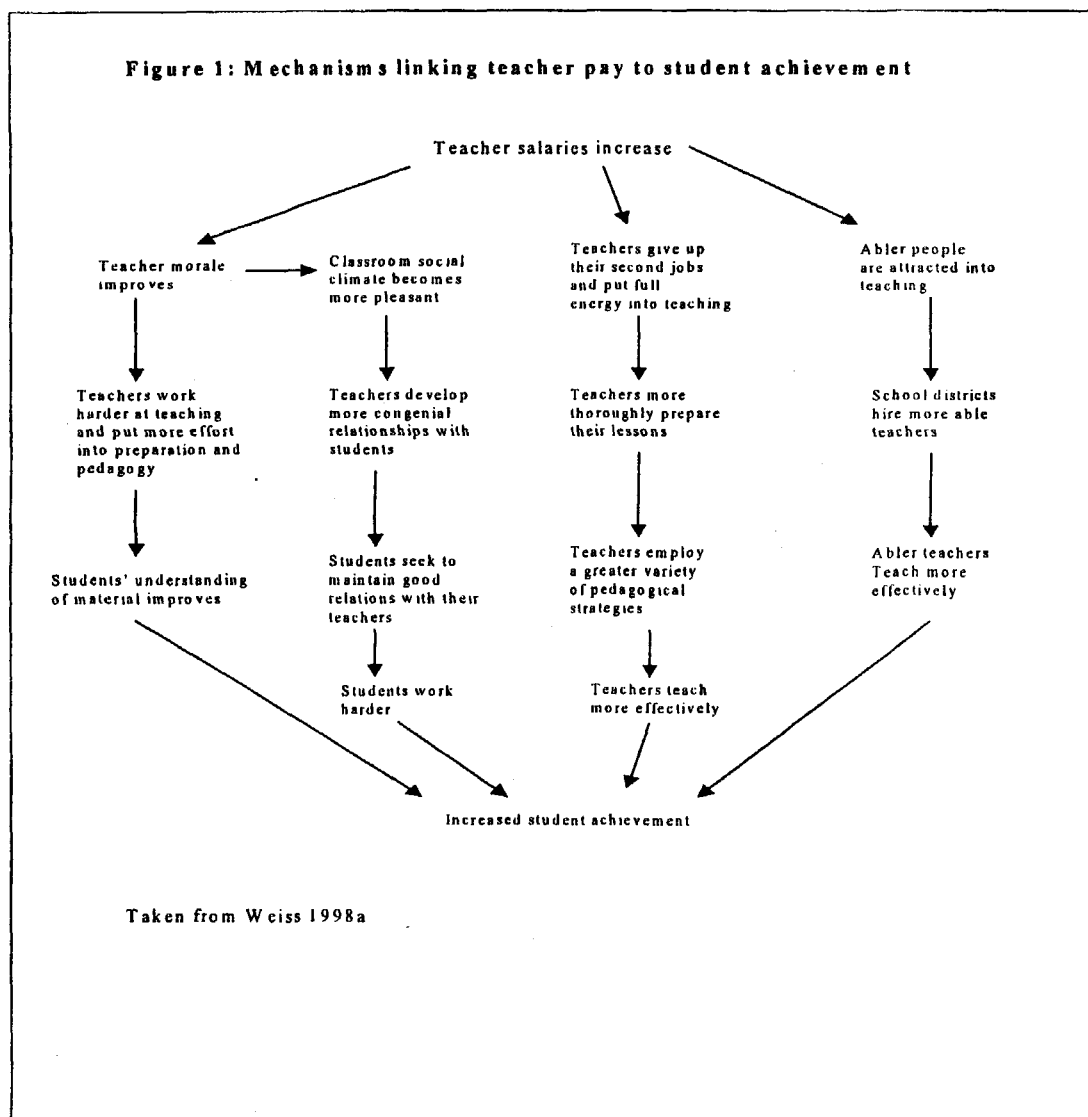
Her decision to emphasize the word "mechanisms" is significant, as we move on to look at scientific realism.

Scientific Realism

Pawson and Tilley have fully described scientific realism in their book *Realistic Evaluation* (1997a). It makes use of four concepts — embeddedness, mechanisms, contexts and outcomes. "Embeddedness" refers to the wider social system within which all human action takes place. Much of it is taken for granted — it is implicit. As an example, they give "the act of signing a check is accepted routinely as payment, only because we take for granted its place within the social organization known as the banking system. If we think of this in causal language, it means that causal powers do not reside in particular objects (checks) or individuals (cashiers), but in the social relations and organizational structures they form" (Pawson and Tilley, 1997b:406). This notion is important to the understanding of precisely what a "program" comprises. It is not, simply, a target hardening initiative or a block watch program: "a program is its personnel, its place, its past and its prospects."

When turning to *how* a program might exert its effect, Pawson and Tilley use the notion of "mechanisms." It is directly analogous to the notion of program theory as described by Weiss and others. Describing the mechanism means going inside the black box. As Pawson and Tilley say, we can never understand how a clock works by

looking at its hands and face, rather we need to go inside the works. It is through the process of understanding, or hypothesizing about mechanisms, that we move from evaluating whether a program works or not, to understanding what it is about a program that makes it work.

Figure 1: Mechanisms linking teacher pay to student achievement

Teacher salaries increase

Teacher morale improves → Classroom social climate becomes more pleasant

Teachers give up their second jobs and put full energy into teaching

Abler people are attracted into teaching

Teachers work harder at teaching and put more effort into preparation and pedagogy

Teachers develop more congenial relationships with students

Teachers more thoroughly prepare their lessons

School districts hire more able teachers

Students' understanding of material improves

Students seek to maintain good relations with their teachers

Teachers employ a greater variety of pedagogical strategies

Abler teachers Teach more effectively

Students work harder

Teachers teach more effectively

Increased student achievement

Taken from Weiss 1998a

But whether it "works" or not will also depend on the "context" within which it is introduced. Solutions to crime problems are often

context-dependent. What might work in one place could be disastrous or prohibitively expensive in another.

As an example, let us look more closely at block watch. If block watch were to reduce crime, how would it do so? What is the "mechanism"? And how is it related to the context within which it may be introduced? The bottom line seems to be that residents agree to "watch out" in their neighbourhood, and call the police if they see anything unusual, particularly an offence in progress. Another way in which it may work would be, through the window decals, to alert would-be offenders to the fact that they are entering an areas where residents care about their neighborhood (whether they do or not) — so offenders better look out! There may, of course, be other reasons for block watch, with other mechanisms coming into play — perhaps it is going to improve police/public relations for example, but that would be a different outcome measure. Let us stick with crime reduction for now.

So we have two hypothesized "mechanisms," and both may be correct and contribute to any observed effect. We also have to ask ourselves to what extent the mechanisms behind block watch may be operating anyway, whether or not we introduce the scheme in a particular area. For example, in low crime middle class areas if residents see a crime in progress they already call the police. So what would we be testing? — perhaps the marginal effect of the window decals and street signs. In this case, we are saying that the mechanism is context-dependent.

So the evaluation of block watch has now become quite complex. We may need a high crime area, where residents would not normally call the police, and we may need to do some very specific things to make sure the residents feel comfortable with doing that — providing telephones for example, and some degree of protection against bullying or threats. We also need to take note of the fact that in areas of this kind it is probably the neighbors who are burgling each other, so looking out for strangers may not be so important as providing a socially acceptable and safe way for residents to call the police. We may also need a low crime area where we are fairly confident that residents do call the police, but we are interested in the effect on burglary of street signs and window stickers — i.e., publicizing the fact to the would-be offender that the neighborhood is cohesive and working against crime. And we may be interested in the effect of publicity alone in high crime areas, regardless of whether the community members sign up to block watch. By setting up a set of inter-related projects, each with a specified mechanism in different contexts, we can begin to see how a body of knowledge can be built up. The picture is set out in Table 1.

Table 1: Mechanisms Operating in Block Watch Crime Reduction Program

Context	Mechanism	
	Call police when crime in progress or when seen a suspicious stranger	**Window/door decals Street signs**
Low crime area	Already operating	New measure introduced
High crime area 1	Special measures taken by program staff to support this	No decals or street signs
High crime area 2	No measures taken to support this	Decals and street signs introduced in the area
High crime area 3	Special measures taken by program staff to support this	Decals and street signs introduced in the area

The mechanism, then, is hypothesized to have a particular effect, in a given context, which will lead to an observed "outcome," in our current discussion, a reduction in crime. This can be summarized as Outcome = mechanism + context, or in prose:

> The basic task of social enquiry is to explain interesting, puzzling, socially significant outcomes (O). Explanation takes the form of positing some underlying mechanism (M) that generates the regularity and thus consists of propositions about how the interplay between structure and agency has constituted the outcome. Within realistic investigation there is also investigation of how the workings of such mechanisms are contingent and conditional, and thus only fired in particular local, historical, or institutional contexts (Pawson and Tilley, 1997b:412).

By operating in this way, and being prepared to replicate our results in other areas, a body of knowledge can be built up which enables the better understanding of the various methods of crime control.

Rival Explanations

Yin presents his ideas for "Rival Explanations" as a methodology for evaluating complex community-based programs (Yin, 1999, 2000), although they are equally appropriate for testing more modest

community or policing initiatives. Yin describes the features of CCIs as typically involving:

- Systemic change (e.g., in the norms, infrastructure or service delivery of an agency or set of agencies), not just change in individual behavior;

- Multi-faceted intervention, not just a single variable;

- Unlikely to be standard across sites;

- Multiple, not single outcomes of interest; and

- Idiosyncrasies of communities reduce validity of defining "comparison" sites.

Yin's argument is that a complex set of activities is introduced into a community; there is a desire to attribute subsequent change to those activities (i.e., to say that they caused them); the threat to this attribution is that some other event or set of events may have been the cause; therefore, let us test plausible rival hypotheses or explanations for the observed changes and see if we can eliminate them. He likens this approach to that or the journalist, detective, forensic scientist or astronomer.

Yin notes that traditional experimental designs try to rule out rival explanations through randomization or the use of control groups, without specifying what the rival explanations might be. Although Yin's rival explanations methods rule out only those rivals that are named and tested for, the process is more transparent. His main argument, however, is that this approach has greater applicability in that it can address more complicated social change, including CCIs.

The steps are fairly straightforward: define the problem; hypothesize about the main effect and any other possible reasons which may explain any observed change (the rival hypotheses); collect evidence, and assess the evidence in support of the main and rival hypotheses. Yin lists a number of possible rival explanations, which may apply in a fairly wide range of community-based evaluations. These are:

- Direct rivals in practice or policy — another intervention which was introduced at the same time *caused* the change;

- Commingled rival — another intervention *contributed* to the change;

- Implementation rival — the initiative was *not implemented* properly and could not have caused the change;

- Rival theory — the change was caused by the intervention but for *other theoretical reasons* than those hypothesized;

- Super rival — some new innovation, but external to the project and applying more generally caused the change; and

- Societal rival — things are changing anyway.

All these various options may need to be considered in any given evaluation, and the program evaluator can never, in theory, rule out every alternative potential cause of change. But by making the alternatives explicit, and collecting data relevant to testing them, the process becomes transparent and challengeable as a result.

Commonalities and Differences

There are some common features to these approaches, which make them attractive. First they unpack the "black box." They require a statement of the causal mechanism between what is being done and what is expected to change. Weiss and her colleagues call this "program theory," Pawson and Tilley call it "mechanisms," and Yin calls it an "explanation." It is one of the most important improvements on the traditional experimental methods approach, and is essential if we are ever to build up a body of knowledge about what works, where and why. It also has other advantages. If, for example, the causal chain is long, or tenuous, then this might cast doubt on the likelihood of the intervention generating a measurable effect, or may lead policy advisers to question its cost-effectiveness.

"Program theory" and "rival explanations" approaches have relatively little to say directly about "context," which is given considerable weight in the discussion of "scientific realism." This concept is important in surfacing the relationship between *what* is being done in a program, which may cause the hypothesized effect, and the *circumstances* or social context within which it is being done. These relationships take on an increased significance when the question of generalizability is raised. It then becomes important to know the extent to which the particular characteristics of the project site were essential to the salience of the mechanism, in order to replicate the results. And this issue becomes even more relevant with the increasing interest of policy advisers in cost-benefit or cost-effectiveness analysis.

To take the Perry Preschool Program (Barnett, 1996) as a concrete example, this involved exposing a group of youngsters at high risk of school failure, involvement in crime and delinquency and so on, in a high quality intensive preschool support program, which included their parents and teachers. The project was highly successful in achieving a number of aims, including less involvement in delinquency, better school performance and reduced teenage pregnancy.

The cost-benefit ratio is quoted as 1:7 — for every dollar spent, seven dollars were saved. So does that mean that the Perry Preschool Program should be implemented nationwide? Would the same benefits accrue? The answer is no. The Perry Preschool Program was introduced in a high crime disadvantaged area of Chicago. The context was right for such a program and the potential payoff was high. Introducing a similar scheme in Bethesda, or Hampstead, both upper middle class areas where the involvement of young people in crime is already low, and teenage pregnancies are not a major social concern, would not be cost-effective. There would be too many false positives in the population — children who were not going to get into trouble whether they were in the (expensive) program or not. Looking at the mechanism through which the Perry Preschool Program was supposed to be effective, and realizing that in many other contexts those mechanisms, or better, are already in place, leads to the realization that mechanisms and context are inter-related.

The rival-explanations approach stands out from program theory and scientific realism in offering an alternative to the "control group" used in traditional experimental paradigms. The program theorists note that it is almost impossible to find a control community with which to compare the CCIs, but they do not say how they cope with the criticism that any observed changes in their program area could have been caused by something that was operating outside the program, perhaps in the whole city or state within which the community was located. Yin's approach to this is highly pragmatic. To take a straightforward example from policing, let us assume that we introduce a crime reduction project in a public housing complex, and we subsequently find a reduction in crime. Was the reduction due to our intervention? Following Yin, we may then look at crime rates in the whole police precinct, or the neighboring areas, to test the rival hypothesis that crime was reducing generally, and for other reasons than our intervention. We would not, in this case, characterize the precinct or neighboring areas as *control* areas with the implication that they are somehow comparable.

In various combinations, these three approaches constitute a viable alternative to the experimental methods traditionally adopted by social science researchers. The extent to which they, and some of the other approaches discussed, meet some of the criteria relevant to delivering what is required of social science research methods is summarized in Table 2 below.

Table 2 requires some explanation. Working across the categories at the top of the table — *"Theory-based'* means that the evaluative approach requires the description of a theory or mechanism through which change will be achieved. Program theory and scientific realist

approaches clearly state the mechanism, but experimental designs do not.

The production of *"timely results"* clearly depends upon the scale of the evaluation. Data can only be collected as fast as they become available. If a program requires the determination of reconviction rates following release from prison, for example, then there may be a built in delay of some years. RCTs are classified as lacking timeliness because the emphasis on the sampling frame, and getting that right, can lead to long delays in establishing the project.

The extent to which project evaluations are *"real-world sensitive"* is an assessment of the extent to which they may make unreasonable or simply undeliverable demands on the practitioners or policy advisers with an interest. For example, the requirements for random allocation can sometimes be quite unacceptable politically, and the need for a program to be "set in concrete" and not to change, develop or evolve in the interests of the evaluators, is in many cases impractical and unrealistic. Experimental methods come out badly here.

The *"user-friendliness"* of an approach does, to some extent, depend on how the results are written up and presented to policy advisers and practitioners. The experimental methods are at greatest risk of being less well understood by non-specialists, since they can often involve fairly sophisticated sampling strategies or statistical analyses.

"Internal validity" relates to the extent to which any changes in the dependent variable — i.e., the thing the program is intended to change (in our case the rate of crime or disorder) — can be unambiguously attributed to the program or project activity. It is here that the experimental designs come to the fore. They are specifically intended to cope with threats to internal validity. Program theory and scientific realist approaches are acceptable, if combined with the rival hypotheses approach.

"External validity" is an assessment of the extent to which the evaluation findings can be generalized to other places, times or populations. Most approaches do not do too well on this criterion, which serves to illustrate the importance of replication as a means of establishing principles of effectiveness, and of teasing out the relationship between contexts and populations.

Table 2 covers a selection of the kinds of criteria that practitioners or policy advisers might be interested in when considering the various types of evaluation, but it does not cover all of their interests, nor does it work systematically through their "needs" more generally, as were discussed above.

Table 2: Research Methods and Delivery of Results

Approach	Theory-based	Timely results	Real-world sensitive	User-friendly	Internal validity	External validity
Program theory	✓	✓	✓	✓	✓	?
Quasi-experimental designs	✗	✓	✗	✗	✓✓	✗
Randomized trials	✗	✗	✗	✗	✓✓	✗
Rival hypotheses	✓	✓	✓	✓	✓	?
Scientific realism	✓	✓	✓	✓	✓	?

Table 3 below summarizes the needs that were identified (without distinguishing between practitioners and policy advisers) and comments on how they might be met.

Table 3: Meeting the Needs of Practitioners and Policy Advisers

Needs	How will the needs be met?
To know what works, where and why	More investment in experiments with an outcome focus and more use of research designs which are explicit on mechanisms and contexts.
Help in replicating "what works" — understanding contexts and mechanisms	Greater preparedness on the part of funding agencies to support replication of apparently effective projects in different contexts. Inclusion of these concepts in training programs for practitioners and researchers.
Cost included in evaluations	Encouragement by central funding agencies to include at least basic cost measures in new projects.
Help in generating testable hypotheses	Closer working relationships between evaluators and practitioners before evaluations are commenced to ensure that the hypotheses being tested are clear and agreed.
Timely research	Funding arrangements, which do not lead to excessive delays in commissioning "hot topics." Tight management of research contracts to minimize delay in completing work and a clear publication or dissemination strategy aimed at getting results out quickly.
Involvement in setting the agenda	Mechanisms in place to ensure that policy advisers and practitioners are involved in the research agenda-setting process.
Reports and recommendations in plain English	Advice to report writers on "house style" and training in report writing where appropriate.
To know of current good practice	Regular reviews of "what works, where and why." Support for the Campbell Collaboration and similar activities. Attention to the means through which good practice is developed and disseminated.

Needs	How will the needs be met?
Feedback on the results of research in which they have participated	Routine mechanism established to ensure that research results are reported back to those who have contributed to their development.
Good news	Sensitive drafting, and early warning of potential bad news. Essentially managing the expectations of the recipients of research results.
Confidence in the results	External scrutiny of research reports; publication of results; transparency; high standards in training and support for researchers.
Risk taking in making inferences from data	Specific encouragement for researchers to explain the implications of their work for policy and practice. This may require them better to understand the constraints under which policy advisers and practitioners operate, which in turn may require specific training opportunities for researchers. A clear statement should also be required of the extent to which any advice is speculative rather than deriving directly from the research data.

Some of the approaches suggested in the table are already operating or are being developed. They are not solely related to the methodology of choice, but where they are relevant to methodology, they are not necessarily compatible with the tradition of experimental design.

CONCLUSIONS

This chapter has argued that the research community has not served the advancement of evidence-based policy and practice for a number of reasons. It has also argued that the current interest in evidence means that the situation has got to change. Not only will the policy advisers and practitioners need to pay more attention to research, but the researchers will need to be more sensitive to the needs of these two groups, who are increasingly being conceptualized

as partners rather than "experimental subjects" in the case of practitioners, or "disinterested funders" in the case of policy advisers.

The kinds of approaches advocated here as more methodologically appropriate pay greater attention to the realities of community crime control. They are not necessarily cheaper, because of the need for replication, but they should lead to a body of knowledge in which we can have greater confidence. They are also addressing the right questions — what works where and how? There are, however, some caveats to the enthusiastic adoption of a closer working relationship between policy advisers and practitioners, which need to be noted. First, attention needs to be paid to the testing of rival hypotheses. The difficulty of identifying control groups for many of the community-based interventions, which are increasingly being proposed as offering the best hope of longer-term crime control, are particularly difficult to evaluate, but this should not rule out the testing of at least the most obvious rival hypotheses.

Secondly there are a number of issues, which can loosely be related to "integrity," which also demand attention. One set of problems can arise if researchers are close to policy advisers, and a related but different set can arise from closer working with practitioners.

It would, of course, be a mistake to assume that researchers using experimental designs are invulnerable to pressure, and that they somehow hold the ethical high ground. There are numerous cases where results have been manipulated and scientific processes ignored during randomized controlled trials — held out as the gold standard in experimental design. Scientists, including "pure" scientists, are as vulnerable to pressures as the next person; what differs is the source of the pressure. In academia, for example, there is pressure to publish research in journals, and this leads to a reinforcement of critical, generally negative findings, and a fiercely competitive context within which research is being carried out; the prize goes to the first in print.

There are essentially two dimensions in this debate: ethics, or integrity and the extent to which researchers and practitioners are working together. Ideally these two dimensions would be orthogonal, but we have no evidence that they are. But neither do we have any evidence of a necessarily high positive correlation, as is implicitly assumed by those who criticise close collaborations of researchers with policy advisers or practitioners.

It is, nevertheless, probably true that proximity to policy advisers risks exposure to pressure to come up with good news, fast. And the notion that researchers might also offer opinions, as is suggested in the last row of Table 3, invites them to go beyond their data. If they were to do this, then they would need to be very clear on the bound-

ary between fact and opinion, both in their own minds and in their communication with their partners.

Working closer with practitioners risks emotional involvement in the results of the research exercise. Sympathy toward hard working and committed colleagues, who are trying their best in difficult circumstances, risks researchers putting a positive spin on the results which is not justified by the data (see Stake, 1997, for an example).

Some options for handling these difficulties have been suggested by Weisburd (1994), whose starting assumption is that the researchers will not pull any punches in their evaluations. He sees the key issue as how best to manage the role tensions that inevitably exist between the practitioners and researchers. Weisburd recommends: carefully mapping out the research design from the outset, with the practitioners fully involved; clear statements made (and agreed) about how emerging information will be handled as the evaluation progresses; a clear definition of what constitutes "success" and how that will be measured; and agreement on how the results will be disseminated and how (if at all) program reports might differ from research reports. This is good advice. The process of working effectively with practitioners (and policy advisers) is as much about managing their expectations as anything else. If they believe that the researchers are there simply to rubber stamp the project efficacy, then they are operating under an illusion and it is important to make the terms of engagement clear. An evaluation is an attempt at an objective assessment of a program or project. Unfortunately this is not always as clear as it might be. As Sherman (2000) remarked at an NIJ seminar: "If you put forward a hypothesis you are presumed to support it." And therein lies the problem.

If, however, we start with the proposition that the researchers are vulnerable to being "soft" on the practitioners — as Scriven (1997), for example, tends to assume — then a rather different set of options emerges. First, the outcome measure needs to be valid, reliable and independent of the evaluators. If the outcome measure is crime-related, for example, then short of actually fiddling the figures, there should be less of a problem than if the study is reporting on a process, or using some other measure, which is dependent upon interpretation. So that is a first step in the corrective direction — choose outcome measures that are valid, reliable and open to independent scrutiny.

The same potential difficulties arise for researchers working close to policy advisers, where there are similar pressures to "prove" that the latest policy initiative works well. Campbell, in his seminal paper "Reforms as Experiments" (1969:409-410), discusses the relationship between the researcher and political expectations of the research:

... specific reforms are advocated as though they were certain to be successful. For this reason knowing outcomes has immediate political implications.... If the political and administrative system has committed itself to the correctness and efficacy of its reforms, it cannot tolerate learning of failure [original emphasis].

In this new, outcome-focused world the policy advisers are in a major bind. On the one hand they still want the "good news" story — the confirmation that the already announced new initiative works — but they also want a valid assessment of the value of the program: *Did* it "work"? Campbell offers a way out of this bind in his 1969 paper, which is as relevant today as it was then, but only marginally more likely to happen. He suggests that we need a shift in political posture away from the advocacy of a specific reform and toward the advocacy of the seriousness of the problem:

> The political stance would become: This is a serious problem. We proposed to initiate Policy A on an experimental basis. If after five years there has been no significant improvement, we will shift to Policy B [1969:410].

Unfortunately, Campbell seems to ignore the probability that politicians do not normally work to five-year time frames. If Policy A is really going to take five years to test, then by the time the results arrive there could well be a quite different administration in place, or at least different individuals, with their own prejudices and concerns. The answer, as with the practitioner, is to be clear at the outset that the results will be what they are, and they may or may not confirm the program as a complete or partial success. If the political position is that regardless of the evaluation results the program will be expanded and declared a winner, then the best advice to the politician is to save the money and not bother with the evaluation in the first place.

Perhaps the most direct and important means of guarding against threats to objectivity is through transparency — the results should be published unless pre-publication peer review suggests that there are methodological flaws which cannot be corrected. The availability of a report on a study should be a requirement regardless of the methodology used, but is less likely to happen if the study is carried out by a consultancy company, not-for-profit organization or professional evaluator with no personal or professional investment in publication. Academics, on the other hand, have a vested interest in ensuring publication. Sufficient detail to enable replication is also important. Publication and scrutiny of results by the academic commu-

nity and by the critical and independent media should help to offset any tendencies toward partiality, exaggeration or significant departures from the "truth."

One final, perhaps overly defensive point: This chapter is not saying there is no place for experimental methods in crime prevention or policing research. But it is stressing the need better to understand the methodological choices and to make these choices on the basis of a thorough understanding of the research questions being addressed, and the policy and practice context within which they are being made. The "gold standard" should not be any *particular* methodology, but a process of informed decision making through which the *appropriate* methodology is chosen.

●

Address correspondence to: Professor Gloria Laycock, Director, The Jill Dando Institute of Crime Science, School of Public Policy, University College London, 29/30 Tavistock Square, London WC1H 9QU. E-mail: <g.laycock@ucl.ac.uk>.

REFERENCES

Barnett, W.S. (1996). *"Lives in the Balance: Age-27 Benefit-Cost Analysis of the High/Scope Perry Pre-School Program."* (Monographs of the High/Scope Educational Research Foundation, #11.) Ypsilanti, MI: High/Scope Press.

Barr, R. and K. Pease (1990). "Crime Placement, Displacement, and Deflection." In: M. Tonry and N. Morris (eds.), *Crime and Justice: A Review of Research,* vol. 12. Chicago, IL: University of Chicago Press.

Bayley, D.H. (1998). "Policing in America: Assessment and Prospects." *Ideas in American Policing.* Washington, DC: Police Foundation.

Campbell, D.T. (1969). "Reforms as Experiments." *American Psychologist* 24:409-429.

Chen, H-T. (1990). *Theory Driven Evaluations.* Newbury Park, CA: Sage Publications.

——and P.H. Rossi (1987). "The Theory Driven Approach to Validity." *Evaluation and Program Planning* 10:95-103.

Christie, N. (1997). "Four Blocks Against Insight: Notes on the Oversocialization of Criminologists." *Theoretical Criminology* 1(1): 13-23.

——(1977). "Conflicts as Property." *British Journal of Criminology* 17(1):1-15

Clarke, R.V. and D. Weisburd (1994).. "Diffusion of Crime Control Benefits: Observations on the Reverse of Displacement." In: R.V. Clarke (ed.), *Crime Prevention Studies,* vol. 2. Monsey, NY: Criminal Justice Press.

Colledge, M., P. Collier and S. Brand (1999). *"Programmes for Offenders: Guidance for Evaluators."* (Crime Reduction Programme Guidance Note 2.) London, UK: Home Office.

Cornish, D. and R.V. Clarke (1987). "Social Science in Government: The Case of the Home Office Research and Planning Unit." In: M. Bulmer (ed.), *Social Science Research and Government* Cambridge, UK: Cambridge University Press.

Crime and Disorder Act, U.K. (1998). "Chapter 37." London, UK: The Stationery Office.

Dhiri, S. and S. Brand (1999). *Analysis of Costs and Benefits: Guidance for Evaluators.* (Crime Reduction Programme Guidance Note 1.) London, UK: Home Office.

Doherty, J. (2000). "Housing: Linking Theory and Practice." In: H.T.O. Davies, S.M. Nutley and P.C. Smith (eds.), *What Works? Evidence Based Policy and Practice in Public Service.* Bristol, UK: The Policy Press.

Felson, M. and R.V. Clarke (1998). *Opportunity Makes the Thief: Practical Theory for Crime Prevention.* (Police Research Series, Paper no. 98.) London, UK: Home Office.

Fulbright-Anderson, K., A.C. Kubish and J.P. Connell (eds.) (1998). *New Approaches to Evaluating Community Initiatives: Volume 2 — Theory, Measurement and Analysis.* Washington, DC: The Aspen Institute.

Goldblatt, P. and C. Lewis (eds.) (1998). *Reducing Offending: An Assessment of Research Evidence on Ways of Dealing with Offending Behaviour.* (Home Office Research Study No. 187.) London, UK: Home Office.

Goldstein, H. (1990). *Problem-oriented Policing.* New York, NY: McGraw Hill.

——(1979). "Improving Policing: A Problem Oriented Approach." *Crime & Delinquency* (April):236-258.

Kennedy, D. (1999). *Research for Problem Solving and the New Collaborations in Viewing Crime and Justice from a Collaborative Perspective: Plenary Papers of the 1998 Conference on Criminal Justice Research and Evaluation.* Washington, DC: U.S. Department of Justice.

Kubish, A.C., C.H. Weiss, L.B. Schorr and P. Connell (1995). "Introduction." In: P. Connell, A.C. Kubish, L.B. Schorr, and C.H. Weiss (eds.), *New Approaches to Evaluating Community Initiatives — Concepts, Methods and Contexts.* Washington, DC: The Aspen Institute.

Laycock, G. (2001). "Hypothesis Based Research: The Repeat Victimisation Story." *Criminal Justice* (London: Sage) l(1):59-82.

———(1996). "Rights, Roles and Responsibilities in the Prevention of Crime." In: T. Bennett (ed.), *Preventing Crime and Disorder — Targeting Strategies and Responsibilities.* University of Cambridge Press.

———and N. Tilley (1995). "Implementing Crime Prevention Programs." In: M. Tonry and D. Farrington (eds.), *Building a Safer Society: Crime and Justice, A Review of Research,* vol. 19. Chicago, IL: University of Chicago Press.

———and N. Tilley (1993). *Policing and Neighbourhood Watch: Strategic Issues.* (Crime Prevention and Detection Paper No. 60.) London, UK: Home Office.

Leigh, A., G. Mundy and R. Tuffin (1999). *Best Value Policing: Making Preparations.* (Police Research Series, Paper 116.) London, UK: Home Office.

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

———and N. Tilley (1997b). "An Introduction to Scientific Realist Evaluation." In: E. Chelimsky and W.R. Shadish (eds.), *Evaluation for the 21st Century.* Thousand Oaks, CA: Sage Publications.

Sherman, L. (2000). Seminar at the National Institute of Justice. Washington DC, April, 2000.

———D. Gottfredson, D. MacKenzie, J. Eck, P. Reuter and S. Bushway (1997). *Preventing Crime: What Works, What Doesn't and What's Promising A Report to the United States Congress.* Washington, DC: U.S. Department of Justice.

Scriven, M. (1997). "Truth and Objectivity in Evaluation." In: E. Chelimsky and W.R. Shadish (eds.), *Evaluation for the 21st Century.* Thousand Oaks, CA: Sage Publications.

Stake, R.E. (1997). "Advocacy in Evaluation: A Necessary Evil?" In: E. Chelimsky and W.R. Shadish (eds.), *Evaluation for the 21st Century.* Thousand Oaks, CA: Sage Publications.

Tilley, N. (1993) *After Kirkholt: Theory, Methods and Results of Replication Evaluations.* (Crime Prevention Unit Paper 47.) London, UK: Home Office.

———and G. Laycock (2000). "Joining Up Research, Policy and Practice About Crime" *Policy Studies* 21 (3):213-227.

U.K. Home Office (1999). *Reducing Crime and Tackling its Causes — A Briefing Note on the Crime Reduction Programme.* London, UK: Home Office.

U.S. National Institute of Justice (1997). *CriminalJustice Research Under the Crime Act— 1995 to 1996.* Washington, DC: U.S. Department of Justice.

Weisburd, D. (1994). "Evaluating Community Policing: Role Tensions Between Practitioners and Evaluators." In: D.P. Rosenbaum (ed.), *The Challenge of Community Policing: Testing the Promises.* Thousand Oaks, CA: Sage.

Weiss, C.H. (1998a). *Evaluation* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

———(1998b). "Have We Learned Anything New About the Use of Evaluation?" *American Journal of Evaluation* 19(l):21-33.

———(1995). "Nothing so Practical as a Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives." In: J.P. Connell, A.C. Kubish, L.B. Schorr and C.H. Weiss (eds.), *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts.* Washington, DC: The Aspen Institute.

———(1972). *Evaluation Research: Methods for Assessing Program Effectiveness.* Englewood Cliffs, NJ: Prentice Hall.

Yin, R.K. (2000). "Using the Rival Explanations Method for Evaluating Complex, Community-based Programs." Paper presented at the Annual Conference on Criminal Justice Research and Evaluation, Washington DC, July 2000.

———(1999) *Rival Explanations as an Alternative to Reforms as Experiments.* Bethesda, MD: Cosmos Corporation.