The Development of Realistic Evaluation Theory through the Evaluation of National Crime Prevention Programmes

Professor Martin Gill, Angela Spriggs, Scarman Centre University of Leicester

This paper explores the theory of realistic evaluation as applied to national crime prevention programmes carried out in the UK.

Evaluations of these programmes have used the CMO methodology developed by Pawson and Tilley (1997). This is developed from their concept of realistic evaluation. It has been used as a framework for investigating aspects of public policy in crime prevention, by identifying ‘transferable lessons’ from crime prevention programmes, that is, aspects of crime prevention programmes which have been found to work and which can be used elsewhere.

This paper will, first, describe our understanding of realistic evaluation as defined by Pawson and Tilley (1997). It will then describe some of the difficulties encountered when the CMO methodology was used in practice. Finally it will show how these difficulties may be overcome and how aspects of the concept behind the CMO methodology may be incorporated into future evaluations to identify ‘transferable lessons’.

What is realistic evaluation (Pawson and Tilley)?

1. The purpose of a realistic evaluation is to establish whether there is an ‘inequivocal causal relationship between a program and its outcome’ (31). That is, where some change can be measured following the installation of a particular programme, it seeks to establish beyond doubt that it was the actual programme which caused the measureable change, and not some other, unidentified variable.

2. Realistic evaluation assumes that there is an underlying theory behind the workings of a particular programme or intervention. This theory explains how the particular programme caused the measured change.

3. The most important aspect of the realistic evaluation is the overall context in which the programme takes place.

4. In order to find the underlying theory, it is necessary to configure a series of CMOs for each intervention. A CMO has three constituent parts: a context, a mechanism and an outcome.

‘The context’ signifies the precise circumstances into which a particular intervention is introduced. The mechanism is the precise way in which this measure works within the given context to produce a particular ‘outcome’. This is illustrated by the following CMO:

<table>
<thead>
<tr>
<th>Context</th>
<th>Mechanism</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is a high burglary rate in a particular street; the burglars do not live on this street.</td>
<td>A Neighbourhood Watch system is developed with visible signs which deter would-be burglars.</td>
<td>The burglary rate decreases on this street.</td>
</tr>
</tbody>
</table>
The precise make-up of the context will greatly influence whether a mechanism can work or not. Using the above example, if the context was changed such that the burglars did live in the same street they would be less likely to be deterred by the signs and the outcome would be that the burglary rate would remain the same. Therefore for a programme to ‘work’ the contextual conditions must be right for allowing mechanisms to 'gel' or 'fire' to produce a defined outcome.

5. Each CMO forms the basis of a ‘mini-experiment’ which needs to be measured empirically.

6. Through measurement of a series of CMOs it should be possible to deduce the features of contexts that allow different mechanisms to work to achieve particular outcomes. Thus ‘transferable lessons’ may be learned.

However, a number of challenges have been faced where the realistic evaluation approach has been used in practice. These will be described.

**Challenges faced in applying Pawson and Tilley’s theory (1997)**

1. The development of meaningful CMOs requires a great deal of skill on the part of the researcher. He or she must have knowledge of criminological theory and previous research findings. Furthermore, he must identify the correct context and deduce how the intervention may interact with that context to produce an outcome.

2. A large number of programmes have been found to be flawed. First, it may suffer from theory failure, where there is a flaw in the initial premise behind the programme, often through misunderstanding of the context in which the scheme has been set up or the theory by which a particular initiative may work. For instance, Neighbourhood Watch schemes could be set up in situations where the burglars live within the area covered.

Secondly, there may be implementation failure, where the scheme is not set up properly. For instance, the Neighbourhood Watch coordinator could resign with no replacement. In this case, the programme may change from one where there is a commitment to Neighbourhood Watch to one where the scheme merely consists of signage. The accompanying mechanisms for these will be fundamentally different.

Given that a large number of programmes fall into these categories, can we write these off in terms of realistic evaluation? What is needed is a way to learn from these, which the CMO configuration does not yet provide.

3. Even where the theory and implementation are not flawed, the programme may evolve over time. For instance, the level of commitment of Neighbourhood Watch participants may change.

These raises a number of concerns:

i. When should the CMO be configured? A CMO configured when the programme is first implemented will be different to that defined at the end of, or during, implementation. Is there any value in configuring different CMOs at different points in the process?
ii. When should a CMO be measured for the purpose of learning transferable lessons? The way in which the programme works or otherwise will change throughout; is there ever a time when it is fully implemented? Should the impact be measured when the Neighbourhood Watch scheme is first set up or when a point of equilibrium is reached? How do we judge this? This may also be artificially determined by the timescale of any particular evaluation.

3. At the last analysis, the absence of a control means that the above provides a method of determining the circumstances in which a particular programme mechanism may or may not have worked, it provides no absolute measure of whether that programme mechanism has worked. There is the capacity for missing the relevant context.

4. An infinite number of CMO configurations can be drawn for the simplest of programmes. Pawson and Tilley have recognised that a number of different contexts and mechanisms can produce the same outcome (p107). Further, they have identified as many as 50 different variables from one initiative. (p108)

This may occur in the following ways:

First, a number of different contexts may be relevant to a given mechanism.

Returning to the CMO example described earlier, an alternative context to the Neighbourhood Watch intervention could be as follows:

<table>
<thead>
<tr>
<th>Context</th>
<th>Mechanism</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is a high burglary rate in a particular street; the burglars do not live on this street; a great deal of fly-posting occurs in the area which means that news signs generally go unnoticed</td>
<td>A Neighbourhood Watch system is developed with visible signs which deter would-be burglars</td>
<td>The burglary rate decreases on this street</td>
</tr>
</tbody>
</table>

In this case, the impact of the particular measure would be reduced.

Furthermore, a particular intervention may work in different ways in the same context. Again if we look at the Neighbourhood Watch example above, an alternative mechanism could be the following:

<table>
<thead>
<tr>
<th>Context</th>
<th>Mechanism</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is a high burglary rate in a particular street; the burglars do not live on this street</td>
<td>The development of NeighbourhoodWatch encourages neighbours to report more incidents</td>
<td>The burglary rate decreases on this street</td>
</tr>
</tbody>
</table>

The sheer number of CMOs creates two main problems:

i. First, there is no a priori method of determining which contexts or mechanisms may or may not be relevant.
Furthermore, each of these individual CMOS forms a mini-experiment in its own right, it is impossible to measure each of these with limited resources.

5. CMOs can be configured at a number of different levels of detail. Pawson and Tilley have identified that hypotheses can be generated at both generalized and specific levels (p116). For instance, the context could be as global as a specific national burglary problem, to which the mechanism could be a national burglary reduction initiative. On the other hand it may be as specific as a burglary on a specific housing estate using a particular modus operandi, or way of committing the offence. Measurement of the global CMO will tell you whether or not the burglary reduction initiative had worked; however it is no better than the experimental approach in deducing why it worked. Conversely, a CMO using a high level of detail will define the context to such a degree that it will be impossible to identify aspects which may be applied in different areas. In other words it will be impossible to identify ‘transferable lessons’. Hence, as Pawson and Tilley identify themselves, it is necessary to aim for a ‘middle ground’.

**So what can we learn from the CMO technique which can be applied to current evaluations?**

1. By its emphasis on context, it prompts the researcher to search for all the contexts which may be relevant to a particular initiative. For instance, in the evaluation of CCTV in which we are currently involved, it has been possible to identify a number of factors, such as the attitude of personnel, the nature of the area, the set-up of the monitoring room, the technical aspects of the camera, which are relevant.

2. The CMO model provides an analytical tool for critically evaluating how particular aspects of the scheme may interact with a particular intervention to produce an outcome. For instance, what precise impact will a specific lens type have on the quality of a CCTV camera image both in daylight and at a particular level of streetlighting? What are the overall crime prevention implications for this? Through this thought process, a CMO has essentially been created; however, unlike the formal realistic evaluation method, no attempt at measurement will be made for each and every one of these: it is only when a pattern is seen to emerge or when the scheme is fully implemented that measurement will be attempted.

3. We have recognised that different CMOs may be configured at different times in the evaluation, which have different implications depending on the intended purpose of the evaluation. Three time-related types of CMOs have been identified. Domain CMOs are configured at the beginning of the programme implementation and establish how the initial programme is likely to work. Range CMOs are configured at the end of the programme and define how the programme actually worked. Layered CMOs track the process between the Domain CMO and the Range CMO.

We are therefore able to suggest the point at which a particular CMO should be measured in order to fulfil a particular evaluation objective, be this policy evaluation, process evaluation or implementation evaluation. It is suggested that the Range CMO defines the level at which transferable lessons can be learned, and is thus the CMO to measured to determine policy. The Domain CMO represents the theory in its purest form. The layered CMO is effectively a process evaluation and may provide lessons for future implementation.
4. Realistic evaluation can be fitted within the quasi-experimental framework to unpack the ‘black box’ more effectively. Thus the experimental framework determines whether something works, and the realistic evaluation determines how it works.

_Bibliography:_