

3 How (and How Not) to Design Research to Inform Policy-making

Ray Pawson and Nicholas Tilley

Imagine, if you will, a senior police officer or policy adviser reading the following conclusion to a review of some twenty publicly funded professional evaluations of the effectiveness of community policing experiments:

The question is, is it (community policing) more than rhetoric? There are ample examples of failed experiments, and huge American cities where the whole concept has gone awry. On the other hand, there is evidence in many evaluations that a public hungry for attention has a great deal to tell the police and are grateful for the opportunity to do so. (Skogan, 1992).

What is the likely reaction to this? What is to be made of it? What has been learnt so far? There are failed experiments and non-failed experiments. So what? Why should a series of evaluation studies be funded at the taxpayer's expense if this is the conclusion? Is the police officer or policy adviser any wit the wiser following a series of studies ending up here?

Our sympathies are entirely with any exasperation felt and robustly expressed. They are emphatically *not* with the social scientists whinnying on with excuses about the technical difficulties and uncertainties of the work. *Nor* are they with the more-research-is-needed chorus, if more research will merely mean more of the same.

Those commissioning evaluations want them so that their policy-making can be better informed. Presumably the reason for bringing in social scientists is to avoid overblown self-serving assessments, which committed practitioners are apt to construct, accounts which are all too liable to collapse at the first critical prod.

We ask three questions in this chapter. First, why do the supposedly most rigorous social science evaluations often produce such desperately disappointing results? Second, is there any alternative to endless

uncertainty and equivocation on the one hand and spurious success stories on the other? Third, can evaluation play a progressive part in taking policy and understanding forward?

We start with the bad news and then move on, in due course, to what we hope is the good.

WHY 'RIGOROUS' SOCIAL SCIENCE EVALUATIONS FAIL

Quasi-experimentation marks the current orthodoxy over high-grade social scientific evaluation. It is shown diagrammatically in Figure 3.1. Its basic logic is desperately simple, and disarmingly familiar to us all. Take two more or less matched groups (if they are really matched through random allocation, you can call it real experimentation; 'quasi-ness' following from the impracticality of this in many cases). Treat one group but not the other. Measure both groups before and after the treatment of the one. Compare the changes in the treated and untreated groups, and, hey presto!, you have a measure of impact. The senior police officer, policy adviser and social scientist are at one in appreciating the beauty of the design. We all are. At one level, it has become an icon of the scientific way of evaluation (Cook and Campbell, 1979), at another it embodies the common-sense reasoning wrapped into advertising campaigns telling us that Washo is superior to Sudz.

It will be necessary to take a brief excursion into the philosophy of the social sciences to explain why quasi-experimentation cannot deliver on its initial promise. However, the point to be made has immediate practical consequences, and to illustrate the argument one of those community policing studies included in the overview, whose conclusion we have

	Pre-test	Treatment applied	Post-test
Experimental group	O_1	X	O_2
Control group	O_1		O_2

FIGURE 3.1 *The classic experimental design*