
INTRODUCTION

by

Johannes Knutsson
Norwegian Police University College

Nick Tilley
University College London

Evaluation has always played a central part in both situational crime prevention (SCP) and problem-oriented policing (POP). In the advance of situational crime prevention, where attempts to prevent crimes are made through opportunity reduction, evaluation has been a driving force. In POP, where tailor-made preventive measures designed to solve problems of crime or disorder are put into effect, evaluation is an integral part of the process, indicating whether the intended effect has or has not come about. There is a growing consensus, however, that much evaluation in crime prevention has been unsatisfactory, but there are also important debates over the best methods of effecting improvements. In what way can the effectiveness of crime preventive measures and programs be assessed systematically and rigorously? What types of evaluation will be most useful to policymakers and practitioners? What have we learned from past evaluations? These questions formed the starting point for the present volume.

The chapters both look backwards to what has been done so far by way of evaluation and forward to what might be done better in the future. The balance between retrospect and prospect varies. The early chapters, especially Knutsson's and Guerette's, focus mainly on what has been done to date. The later chapters, especially Henry's and Johnson's, focus mainly on what might fruitfully be added in the future. In varying ways, the

chapters in between the aforementioned ones reflect on prevailing forms of evaluation and how they and their use might be improved (Braga & Bond; Eck & Madensen; Tilley).

The number of evaluations of situational crime prevention initiatives has grown very substantially since that approach was established in the late 1970s (Guerette). These programs use a variety of methods, but seem to provide compelling evidence of the effectiveness of prevention measures attuned to specific crime problems. A potential problem in discussing suites of past evaluations is, of course, publication bias – the well-known tendency for positive findings to be published rather than negative ones. Though this may, of course, be present in the case of situational crime prevention, it would appear more likely that the masses of situational measures routinely taken, for instance by businesses, simply remain unstudied and unreported. Moreover, the reasons for this have less to do with lack of resources or technical difficulties than with the self-evidence of success for these methods among businesses attempting to reduce their crime losses. Take, for example, the decision of a small retailer in a high crime neighbourhood to place high-value items behind the counter rather than on open display. This is clearly a situational measure, clearly makes the theft of those items practically impossible, and clearly in that sense deals effectively with that businesses's problem. The business itself is unlikely to be concerned with potential displacement of the offense to other merchants.

The very pervasiveness of situational means of preventing crime, and the lack of interest in them when they appear to be helping prevent crime – for example, requirements for employees to submit receipts to claim expenses; locking houses and cars to inhibit theft; not leaving valuables about where they might be picked up by strangers; having guards in art galleries and museums to deter thieves; producing banknotes whose forgery is difficult; and attaching dye-tags to high-value items aimed at youngsters in clothes shops so that their value is reduced if they are stolen – are seldom if ever formally evaluated because there is no interest in conducting the evaluation. Moreover, where evaluations are undertaken they can be carried out where measures are put in place unreflectively or where there is no relevant problem to address. A case in point is closed-circuit television (CCTV) in the U.K., for which there has been massive funding following upon widespread political and public support. This has led to widespread deployment of CCTV without the prior analysis of need or potential that is called for in situational crime prevention. Negative findings about effectiveness here may reflect mindless application rather than weaknesses

in situational crime prevention as a whole or in the well-targeted use of CCTV in particular (Guerette). There is no reason to suppose that evaluations assess a representative sample of situational crime prevention interventions. There are, thus, various sources of selection bias affecting what is evaluated. On balance, these are likely to overlook the obviously effective measures and to home in on the more dubious approaches.

Situational crime prevention evaluations adopt a range of methods and these are reflected in the papers collected here. Though randomised control trials (or their close counterparts) may sometimes be possible (Braga & Bond; Guerette; Henry), many of the papers suggest that they are not always either necessary or ideal (Eck & Madensen; Henry; Tilley; Johnson). Particular problems relate to the fact that situational crime prevention initiatives tend often to be area-based, where the conditions needed for randomised control trials are especially difficult to create, though there are some examples where it is claimed that they are met satisfactorily (see Boruch et al., 2004). Randomized Controlled Trials (RCTs) – or even their close cousins, where some of the conventional conditions are breached – may sometimes be impossible, may sometimes yield invalid findings, and may sometimes be an inefficient methodology for advancing situational crime prevention theory, policy and practice. Other approaches may, thus, be more practicable or preferable in producing high quality findings. Suggestions include simulations (Johnson), examination of detailed expected outcome footprints (Eck & Madensen; Tilley), and regression discontinuity designs (Henry). There is certainly no reason to confine attention to evaluations that can be done or have been done using RCTs or their close equivalents, an approach that is emphasised in the Campbell Collaboration (see below). This book brings out well the fact that much has been and can continue to be learned from evaluation studies using different designs. This is not, though, to deny a place for RCTs when they are practicable and when critical discussion suggests that they will produce the most useful and valid findings for theory, policy and practice.

The papers brought together here bring out the practical difficulties that are faced in producing strong evaluation studies. Many situational crime prevention evaluations are methodologically inadequate, whatever particular approach is adopted (Knutsson). Where local evaluations are undertaken, for example of problem-oriented policing initiatives, this is especially liable to be the case (Knutsson). The schemes are often small and complicated and data are weak. And even where analysts are employed

by the sponsoring agency, they are often not used in the evaluations and may lack the particular skills needed to produce findings that are capable of withstanding critical scrutiny (Knutsson). Moreover, crime prevention initiatives frequently encounter more or less serious implementation problems (Knutsson; Tilley). They are also characteristically introduced in complex settings where much else is going on in addition to the crime prevention activity, and in which offender, practitioner and policymaker stakeholders are liable to adapt their behaviour to one another (Henry; Tilley). Such implementation weaknesses and inherent complexity produce substantial challenges for the production of robust and useful evaluation findings. Thus, although as Guerette shows there are by now numerous studies that show the effectiveness of situational crime prevention initiatives, producing strong studies is more fraught with difficulties than might at first sight seem to be the case.

Evaluations serve varying purposes. Evaluations may be conducted in the interests of informing policy decisions: e.g., where should resources be allocated and which policies should be disseminated? (See mainly Braga & Bond; Johnson.) They may also be conducted to inform practitioners' and policymakers' decisions when faced with a new situation: e.g., what should we do to react to that change? (See mainly Guerette; Knutsson.) They may also be conducted in the interest of taking a field of applied knowledge forward: e.g., how do we develop, refine and test theory? (See mainly Eck & Madensen; Henry; Tilley.) The differing purposes of evaluation clearly overlap, but they may suggest the different standards, priorities and methods that are emphasised in the chapters that follow in this volume.

The problem of bias is clearly important in evaluation, and it can creep in, in many different ways. We have already mentioned some sources. Others that are discussed in this volume relate to temptations to accentuate the positive and to ignore the negative (Knutsson; Tilley). These temptations can come from scholars themselves, whose preferred theories or ideology may lead them to be selective in their analyses and in their choice of findings to report. The temptations can also derive from external pressures. In police services and government departments, for example, it will often be clear that those asking for the evaluation hope for and sometimes confidently expect positive findings. Both junior researchers within these organisations and external contractors can often be drawn into trying to present findings in the most favourable light possible.

It seems to us that there is a risk that experiments, in the sense of RCTs and their closest counterparts that compare apparently similar treatment and non-treatment cases or groups, will take on an unwarranted

status as the unequivocal “gold standard” in evaluations. More seriously, they may thereby also be treated by potential users as well as members of the research community as the only methods that can produce dependable findings. Against this exclusivity, in addition to the points that emerge from many of the papers published in this volume, there is by now a massive and growing literature, both within and beyond the field of situational crime prevention, which shows this view to be untenable (see, for example, Cartwright, 2007; Berwick, 2007; Clarke and Cornish, 1972; Hollin, 2008; Pawson, 2006; Pawson and Tilley, 1997). It is likely that only the most rabid experimental enthusiasts would want to rule out other evaluation designs. Most, if not all, serious practitioners and theorists of experimentalism acknowledge that other methods are sometimes necessary in practice and can yield valid findings.

The Campbell Collaboration is the organisation most associated with experimentalism. Yet Donald Campbell, after whom the movement is named, stressed, “the obvious theoretical complexity of measures and the frequently plausible argument that the measured change was the result of irrelevant components of the complex” (1999, p. 167). He favoured “exploring multiple approaches to measuring ‘the same thing’ ” (ibid.). He further averred that “all available measures have less than perfect validity, being systematically biased” (ibid.). His advice to “governmental science policy makers” was:

Give up the notion of a single new evaluation designed to support a single administrative decision regarding expanding or curtailing a program. Substitute for this the development of a disputatious mutually monitoring, applied scientific community that will advise governmental decisions on specific programs from its general wisdom about research in the problem area. (Campbell, 1999, p. 169)

We agree. Multiple evaluations using multiple measurement methods and multiple designs with mutual checking and criticism are, we think, the approach that is most likely to contribute not only to well-reasoned advice, but also to progress in the fields of situational crime prevention and problem-oriented policing.

REFERENCES

- Berwick, D. (2007). The science of improvement. *Journal of the American Medical Association*, 299, 1182–1184.
- Boruch, R., May, H., Turner, H., Lavenberg, J., Petrosino, A., de Moya, D., Grimshaw, J., & Foley, E. (2004). Estimating the effects of interventions that are deployed in many places. *American Behavioral Scientist* 47, 608–33.

- Campbell, D. (1999). Sociological epistemology. In D. Campbell & M. Jean Russo (Eds.), *Social Experimentation*. Thousand Oaks, CA: Sage.
- Cartwright, N. (2007). *Hunting causes and using them*. Cambridge, UK: Cambridge University Press.
- Clarke, R., & Cornish, D. (1972). *The controlled trial in institutional research: Paradigm or pitfall for penal evaluators*. Home Office Research Study 15. London: Her Majesty's Stationery Office.
- Hollin, C. (2008). Evaluating offending behaviour programmes: Does randomization glister? *Criminology and Criminal Justice* 8(1), 89–106.
- Pawson, R. (2006). *Evidence-based policy*. London: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.