

BOOK REVIEW

Evaluation for the real world: The impact of evidence in policy making

Colin Palfrey, Paul Thomas and Ceri Phillips
Bristol: The Policy Press, 2012, 240 pages, £24.99 (pb)
ISBN 978 1 84742 914 8

Elanor Warwick*
King's College, London

The demise of the English Audit Commission in 2010 ended a period of explosive growth for centralised Government auditing and performance measurement of public sector agencies. *'Evaluation for the real world'*, written during the subsequent transfer of the evaluation process to the voluntary and private sector, is a timely reminder of the contribution of evaluative theory to evidence-based policy-making in Britain during the 1990s and 2000s. And in the current climate of conviction-led policy, it is a reminder of the role of evaluation as a professional, an academic, and above all a political activity.

The book traces the history of evaluation since the mid-20th Century, with chapters covering evaluation methodology in all its varied forms: from formal audits, through naturalistic, realistic or theory-driven approaches and finally to goal-free evaluation. It provides a careful, comprehensive and detailed history of the field, cataloguing theories, listing key authoritative papers and analytically assessing the intellectual careers of key evaluation academics.

Evaluation as a discipline, as theory or practice?

Targeted at a mixed audience of policy and public sector management students, academics, and practitioners working across all these fields, the authors are careful to distinguish between practical evaluation and evaluation research. The former covers fifty years of evaluation undertaken by professional practitioners; the latter is the basis for more recent attempts to raise the status of evaluation as an academic discipline. This endeavour to add an academic gloss to the professional activity of evaluation, and the growing body of theory and literature justifying it as worthy of intellectual and academic concern, colours the tone of the book and raises the inherent tension between "scientific" and pragmatic evaluation. The scientific search for evidence, floundering under the burden of proof against a null hypothesis, is compared to practitioners pragmatically (cynically?) allocating resources to confirm the effectiveness of the programmes they are running. This can lead to 'policy-led evidence making', with

practitioners rarely progressing beyond superficial output evaluation. That is, they merely assess whether a policy output occurred, rather than providing a detailed analysis of the process and focusing on the context, the circumstances and the mechanisms that stimulated change.

The authors maintain that the traditional academic focus on experimental-led evaluation has been discredited; its positivist scientific stance replaced by the relativist ontology and epistemology of Lincoln and Gubas' 'fourth generation evaluation'. This latest stage in evaluation theory aims to reflect the broader claims, concerns and issues of stakeholder audiences engaged in a collaborative and participatory process. In doing so it seeks to unite the objective approaches of professional evaluators (applying managerial tools such as PQASSO or ISO 9001) with the interpretative complexities of social research.

Evaluation as unpopular truth telling

Yet the authors argue that evaluation cannot be purely objective, meaning the authenticity of the process is what matters most. When evaluation is unavoidable, it must be undertaken fairly, result in real impacts, and be openly assessed on its effectiveness. "*Evaluation, because it is a form of social research, is expected to produce findings based on a **fair-minded representation of facts and opinions***" (p.43 - emphasis added).

The central sections describe the methods for gathering evaluation data, and the opportunities for and philosophical, methodological and operational limitations to the various evaluation mechanisms and models. They identify a dozen commonly used evaluation criteria and usefully catalogue the vocabulary and concepts used. At times these explanations seem too familiar and simplistic, more suitable for a glossary (e.g. listing the differences between formative or summative evaluations, induction or deduction, or outlining the process of scientific enquiry). However, an explanation for this definitional precision is given towards the end of the book. While acknowledging that writing solely for an academic audience is a legitimate aim, the authors recognise the risk of over-crafting impenetrable phrases; gently chiding the academics they cite for their inaccessible scholarly style and unnecessarily obscure phrases. They highlight the danger of the 'lexicon of specialist language' becoming a barrier to communication with more practical audiences. This is yet another example of the unequal power relationship between the evaluator and those being evaluated.

The non-utilisation of evaluations

Despite its title, the text is more concerned with evaluation than evidence; and more with evaluation as a theoretical process than as a source of evidence for active decision-making. The authors decry the lack of post-evaluation research, reinforcing the fact that evaluation is not a complete end in itself. One area requiring further investigation is the (lack of) impact of evaluation on Government policies in the UK. The final couple of chapters speculate why recent Governments of varied political persuasions may not have acted on the evaluations they commissioned. In a period where policy makers' first thoughts are the cost of public policies and services, rather than their impact, evaluation "*is a servant and not an equal of politicians*" (p.29). The chapter on economic evaluation illustrates the extent to which public sector performance evaluation has been conflated with the political quest for effectiveness, efficiency and economy.

Many explanations for evaluations being overlooked are provided. Findings may be ignored for a range of reasons; lack of resources, poor methodology or evaluation technique, weak project management or timing, poor communication of the findings, lack of succinct and pithy recommendations, political or ideological motivation, drift away from the original policy question, organisational conflicts, vested interests, even the perception of those being evaluated that the evaluation constitutes a threat to their ongoing project. This last situation results in an entertaining typology of 'pseudo-evaluations' devised by Suchman (1967 cited on p.168), who distinguishes superficial and shallow 'eyewash' evaluations from 'whitewash' ones intended to cover up programme failures. He also identifies 'submarine' evaluations undertaken with the predetermined aim of undermining and sinking a project; and those which provide merely posturing lip service to the process as a diversion to postpone any practical action.

The recent establishment of the 'What Works Centres' in the summer 2013¹ may signal the UK Government's renewed interest in the role of evidence and research in policymaking, but one questions whether this will extend to reinstating a robust evaluation of policy impact? The book maintains a cynical/realist position, referring to the *angst* arising within the field at the failures to learn from the intelligent scrutiny of the evidence that good evaluation can provide. This pessimistic tone recalls Flyvbjerg's *cri de coeur* for committed and transparent evaluation. Describing the application of sustainable town planning policy in Aalborg Denmark, he accuses the associated policy evaluation of irrelevancy. "*The result seems predetermined, and the evaluations... become more ritual than real*" (Flyvbjerg, 1998, p.18). He sees power and politics overwhelming any rational objective assessment of the planning options and ultimately obscuring the impact of the policies.

The real world that Palfrey, Thomas and Philips describe is one occupied by academics, professional evaluators, politicians, policy makers, and of course those being evaluated. Each will view evaluation differently along a spectrum ranging from a technical data gathering process to a tactical activity, in turn, giving differing weight to the interpretations or explanations provided. However, if evaluation is to rise above ritual, it needs more than theoretical or methodological rigour. A far more prosaic, practical circumstance is required. The book admirably highlights the many tortuous barriers that constrain the application of evaluation findings to improve public policy and resultant services, but as the authors succinctly conclude "*evaluations should not only be useful, but actually be used*" (p.215). However informative the findings of evaluation may be, changes to public policy ultimately remain in the gift of those commissioning the evaluation, rarely those undertaking them.

* Correspondence address: Elanor Warwick, Department of Geography, King's College, The Strand, London WC2R 2LS. Email: elanor.warwick@kcl.ac.uk

Notes

1. <https://www.gov.uk/what-works-network#core-functions-of-the-what-works-evidence-centres>

p. 92. Book Review: Evaluation for the real world: The impact of evidence in policy making

References

Flyvbjerg, B. (1998) *Rationality and power: Democracy in practice*. Chicago and London: The University of Chicago Press.

Suchman, E.A. (1967) *Evaluative research*. New York: Russell Sage.