



The Magenta Book

Guidance Notes for Policy Evaluation and Analysis

Chapter 1: What is Policy Evaluation?

Published: July 2003

Government Chief Social Researcher's Office
Prime Minister's Strategy Unit
Cabinet Office
Admiralty Arch
The Mall
London SW1A 2WH

Contact: phil.davies@cabinet-office.x.gsi.gov.uk
Telephone: 020 7272 1862



Chapter 1

What is Policy Evaluation?

Background Paper

The Demand for Policy Evaluation

The need for good analysis and sound evaluation to be at the heart of policy making has been recognised in a number of government publications (Cabinet Office, 1999a, 1999b, 2000, 2001). The *Adding-It-Up* Report (Cabinet Office 2000), for instance, argued that:

Rigorous analysis and, where appropriate, modelling is in the best interests of both Ministers and senior officials. They lead to better decisions and improved policy outcomes. Without soundly based analysis and modelling, those involved in the formulation of policy and the delivery of services will work in the dark. As a result, the pace of reform may be slow.

(Cabinet Office 2000: 3)

Some guidance on evaluation and appraisal is already available from within Government. HM Treasury produces a guide on economic appraisal and analysis, known as *The Green Book*. This distinguishes between *ex ante* appraisal of policy options and *ex post* evaluation of policies that have been implemented. *The Green Book* is mostly concerned with evaluating policies, programmes and projects using economic appraisal techniques. Policy evaluation across government, however, has a wider meaning and uses a variety of analytical tools and methodological procedures from a wide range of academic disciplines. This is the focus of *The Magenta Book*.

Other Sources of Guidance on Policy Evaluation

Other guidance on evaluation from within government comes from The Office of the Deputy Prime Minister (ODPM - formerly the Department of Transport, Local Government and the Regions – (DTLR)), and the Regulatory Impact Unit. ODPM/DTLR has published an *Integrated Policy Appraisal Framework* (DTLR, 2002) that goes beyond appraising the economic consequences of policies, programmes and projects. It provides “a good practice tool designed to help assess the impact of policy proposals” in terms of economic, social, environmental and distributional consequences, as well as risk assessment. ODPM has also produced guidance “for the *ex ante* and *ex post* assessment of expenditure projects, programmes and policies with a spatial and distributional focus” (ODPM/*EGRUP*, 2003:3). The

major focus of *EGRUP*'s guidance is on “policies and programmes aimed at urban and rural regeneration, community renewal and regional development.”

The ODPM/DTLR *Integrated Policy Appraisal Framework* complements the appraisal requirements of the Regulatory Impact Unit (RIU), which “assesses the impact, in terms of costs, benefits and risks, of any proposed regulation which could affect businesses, charities or the voluntary sector” (RIU, 2003). This guidance sets out when policy makers are required to carry out a Regulatory Impact Assessment, and how they should do so. It also incorporates guidance on handling European proposals, previously covered in the Guide to Better European Regulation, and updated to reflect current best practice. The guidance gives examples of good practice and places increased emphasis on wider economic, social and environmental impacts.

The Magenta Book complements all of these sources by providing guidance for social researchers, other analysts, and policy makers on the wide range of evaluation methods used in policy evaluation.

What is Evaluation?

Evaluation has been defined as a family of research methods which seeks “to systematically investigate the effectiveness of social interventions....in ways that improve social conditions” (Rossi, Freeman and Lipsey, 1999:20). Another definition of evaluation is “the process of determining the merit, worth, or value of something, or the product of that process” (Scriven, 1991). Drawing upon these two sources the following definition of policy evaluation can be proposed:

“Policy evaluation uses a range of research methods to systematically investigate the effectiveness of policy interventions, implementation and processes, and to determine their merit, worth, or value in terms of improving the social and economic conditions of different stakeholders.”

The importance of a *range* of research methods is paramount. Policy evaluation uses quantitative and qualitative methods, experimental and non-experimental designs, descriptive and experiential methods, theory based approaches, research synthesis methods, and economic evaluation methods. It privileges no single method of inquiry

and acknowledges the complementary potential of different research methods. The methods used in policy evaluation and analysis are usually driven by the substantive issues at hand rather than *a priori* preferences (Greene, Benjamin and Goodyear, 2001).

What Types of Evaluation Are Used in Government?

Summative and Formative Evaluation

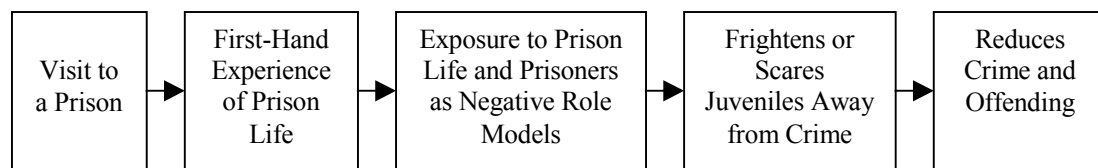
Two types of evaluation that are commonly used in government are *summative* and *formative* evaluation. Summative evaluation, which is sometimes referred to as *impact evaluation*, asks questions such as: What impact, if any, does a policy, programme or some other type of government intervention have in terms of specific outcomes for different groups of people? It seeks to provide *estimates of the effects* of a policy either in terms of what was expected of it at the outset, or compared with some other intervention, or with doing nothing at all (i.e. the counterfactual).

Formative evaluation, which is sometimes referred to as *process* evaluation, asks *how, why, and under what conditions* does a policy intervention (or a programme, or a project) work, or fail to work? These questions are important in determining the effective development (i.e. formation), implementation and delivery of policies, programmes or projects. Formative evaluation typically seeks information on the *contextual* factors, mechanisms and processes underlying a policy's success or failure. This often involves addressing questions such as *for whom* a policy has worked (or not worked), and why.

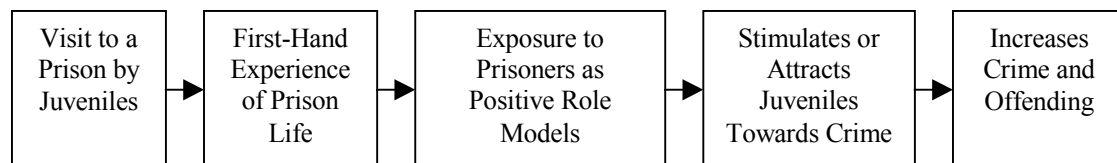
This distinction between summative and formative evaluations is not always as rigid as the above characterisation might suggest. Proponents of the *Theories of Change* approach to evaluation (Chen, 1990; Connell *et al*, 1995; Funnel, 1997, Owen and Rodgers, 1999; Weiss, 1997; Judge and Bauld, 2001) would argue that determining whether or not a policy has worked, or has been effective, necessarily involves asking questions about *how* it has worked, *for whom*, *why*, and *under what conditions* it has worked or not worked. Nonetheless, the contrast between evaluating *whether* a policy intervention has been effective (summative evaluation), and *why* it has done so (formative evaluation), is one that is conventionally made in the policy evaluation literature.

What is Theory-Based Evaluation?

Theory-Based approaches to evaluation, which include the *Theories of Change* approach mentioned above, as well as *Programme Theory Evaluation* (Rogers *et al*, 2000) and some aspects of *Realistic Evaluation* (Pawson and Tilley, 1997), focus on unpacking the theoretical or logical sequence by which a policy intervention is expected to bring about its desired effects. Theory-Based approaches attempt to identify the *mechanisms* by which policies and/or programmes might produce their effects. For instance, the common underlying theory of the juvenile awareness programmes for preventing juvenile delinquency (such as the ‘Scared Straight’ programmes in the United States, (Petrosino, Turpin-Petrossino, and Buehler, 2002)) suggest the following sequential steps:



An alternative possible sequence of outcomes, which can be tested empirically, might be as follows:



Failure to be clear about the causal sequence by which a policy is expected to work can result in well intentioned policies being misplaced, and outcomes that are contrary to those that were anticipated. Theory-Based evaluation provides a number of ways of carrying out an analysis of the logical or theoretical consequences of a policy, and

can increase the likelihood of the desired outcome being achieved. Theory-Based initiatives will be considered in greater detail in Chapter 7.

Can the Policy, Programme or Project Be Evaluated?

Another important question to ask is whether or not a policy, programme or project can be evaluated at all. Some policy initiatives and programmes can be so complicated and diffuse that they have little prospect of meeting the central requirements of evaluability. These are that the interventions, and the target population, are clear and identifiable; that the outcomes are clear, specific and measurable; and that an appropriate evaluation design can be implemented (Patton, 1990).

Have the Goals of a Policy, Programme or Project Been Achieved?

This is one of the most frequently asked questions in policy evaluation, and is sometimes referred to as Goals-Based evaluation. In the American evaluation literature it is sometimes referred to as ‘legislative monitoring’, because it monitors whether the outcomes that were expected from some government policy initiative have been achieved. In the United Kingdom, the achievement of targets that have been set by Public Service Agreements and Service Delivery Agreements are evaluated using Goals-Based methods of evaluation.

An example in the UK context might be whether or not the goals and targets of the National Literacy Strategy (i.e. increasing the reading, writing and comprehension abilities of children and adults) have been achieved. Another example might be whether the goals of the Hospital Waiting Lists initiative (i.e. reducing the number of people on hospital waiting lists and/or the time they had to wait for treatment) have been achieved. Such outcomes may, or may not, be made explicit in policy statements and documents.

Goals Based evaluations make no assumptions about whether or not the chosen goals or targets are valid or appropriate measures of effectiveness. It may indeed be the case that waiting no more than four hours for hospital treatment is less valid to patients and their carers than waiting for two hours or less. Or it may be that waiting times for treatment are less valid than making sure that the most

effective and evidence-based treatment methods are used by doctors and hospitals. Goals Based evaluations simply measure whether some goals or targets set by policy makers have been achieved.

Even when the goals of a policy, programme or project have been achieved, however, this does not necessarily mean that the policy in question has been responsible for this outcome. Other factors, including other policy initiatives, may have been responsible. In order to know whether the policy in question has been responsible for an anticipated outcome, some evaluation of the *counterfactual* is required (i.e. what would have happened anyway, or because of other interventions). Randomised control trial methods are generally considered to be the most appropriate way of determining the counterfactual of a policy, programme or project, though carefully controlled matched comparisons studies and some forms of statistical modelling also provide estimates of the counterfactual. These methods will be reviewed in Chapters 6, 9 and 10 of *The Magenta Book*, and further guidance on how statistical modelling can estimate the counterfactual can be found in *The Green Book* (www.hm-treasury.gov.uk/greenbook)

How Do You Evaluate Unintended Outcomes?

Policy makers and evaluators are often interested in the *unintended* consequences or outcomes of a policy, programme or project. These unintended outcomes may be beneficial or harmful. *Goals-free* evaluation does this by focusing on the *actual* effects or outcomes of some policy, programme or project, without necessarily knowing what the intended goals might be. This type of policy evaluation is more commonly undertaken by evaluators who are independent of government and who are more interested in the *range* of consequences of a policy, programme or project than in the anticipated outcomes alone. Goals-free policy evaluation, however, should be of interest to government social researchers and policy analysts because of the importance of establishing the balance between the positive and negative consequences of policies. Such a balanced evaluation is important in order to establish the cost-benefit and cost-utility of a policy or programme intervention.

What is Experimental and Quasi-Experimental Evaluation?

Experimental and quasi-experimental research methods provide valid and reliable evidence about the *relative effectiveness* of a policy intervention compared with other policy interventions, or doing nothing at all (sometimes called the *counterfactual*). They provide appropriate evidence about questions such as whether a personal adviser service is more, or less, effective in terms of advancing low paid people in the labour market than, for example, providing skills training, or doing nothing at all.

The purest form of experimental method is the *randomised controlled trial* (RCT - sometimes called the *random allocation method* of evaluation). Randomised control trials deal with the problem of other possible factors influencing an outcome by exposing an experimental group of people, and a non-experimental (i.e. control) group of people to exactly the same factors *except* the policy, programme or project under investigation. The allocation of people to the experimental policy intervention, or to the control (i.e. no intervention) situation, is done purely on the basis of chance (i.e. randomisation). Randomisation does not guarantee that the experimental and control groups will be identical, but it reduces the influence of extraneous factors by ensuring that the only differences between the two groups will be those that arise by chance.

Randomisation may be by individuals or by units, clusters, or whole areas. Some welfare-to-work initiatives have allocated individuals (e.g. Job Seekers' Allowance claimants, people on Working Families Tax Credits) to experimental or control groups. Other policy initiatives have allocated units such as schools, hospitals, housing estates or entire neighbourhoods, to experimental or control groups. The methods, problems and limitations of randomised controlled trials are discussed in Chapter 6.

Quasi-experimental methods refer to those research designs that compare the outcomes of experimental and control groups by methods other than randomisation. These include:

- controlled before and after designs involving pre-test and post-test comparisons using a *single group* of people (i.e. where individuals or units are their own controls).

- controlled before and after designs in which pre-test and post-test comparisons are made between *two or more groups of people* (i.e. experimental and external controls).
- interrupted time series studies (based on repeated observations over time of valid and reliable standardised measures of outcome).
- various types of matching designs using matched comparisons of individuals or units before and after an intervention.
- regression discontinuity designs

Experimental and quasi-experimental designs are discussed in greater detail in Chapter 6.

What are Qualitative Evaluations?

Qualitative evaluations are designed to “permit the evaluator to study selected issues in depth and detail” Patton (1990). Such depth and detail is usually necessary to determine the appropriate questions to ask in an evaluation, and to identify the situational and contextual conditions under which a policy, programme or project works or fails to work.

Qualitative methods of evaluation are particularly important for formative evaluation which, as Patton (1990:156) again suggests, “is limited entirely to a focus on a specific context”. Patton goes on to argue that:

Formative evaluation services the purpose of improving a specific program, policy, group of staff (in a personnel evaluation), or product. Formative evaluations aim at ‘forming’ the thing being studied....There is no attempt in formative evaluation to generalise findings beyond the setting in which one is working. The purpose of the research is to improve effectiveness within that setting.

(Patton, 1990, 156)

Qualitative evaluation uses a range of methods including in-depth interviews, case studies, consultative methods, focus groups, ethnography, observational and participant-observational studies, and conversation and discourse analysis. These methods of qualitative evaluation are discussed in greater detail in Chapter 8.

What is Economic Appraisal and Evaluation?

Policies, programmes and projects involve the allocation of scarce and finite resources to competing demands and interests. The old adage that a pound cannot be spent twice means that choices between the competing demands upon a resource have to be made. Consequently, it is necessary to undertake economic appraisal at the outset (i.e. *ex ante*) of different policy options and the likely outcomes (both positive and negative) that will be achieved by them, and of the costs involved in achieving these outcomes. It is also necessary to undertake an economic evaluation after (i.e. *post hoc*) a chosen policy, programme and project has been running for some time in order to determine whether or not the anticipated outcomes (or other outcomes) have been achieved.

There are different types of economic appraisal and evaluation. The simplest type is cost appraisal and evaluation, which simply compares the costs of different initiatives without considering the outcomes to be achieved (or that have been achieved). The limitations of such appraisals and evaluations are fairly obvious – they tell us very little about the *relative effectiveness or benefits* of different interventions – and are of little value alone in policy evaluation.

Other types of economic appraisal and evaluation, which are more analytically powerful and useful to policy making, include *cost-effectiveness* and *cost-benefit* analyses. The former compares the differential costs involved in achieving a given objective, whereas the latter considers the differential benefits that can be gained by a given expenditure of resources. Cost benefit analysis involves a consideration of alternative uses of a given resource, or the *opportunity cost* of doing something compared with doing something else. Another type of economic appraisal is *cost utility* analysis, which evaluates the utility of different outcomes for different users or consumers of a policy or service. Cost utility analysis typically involves subjective appraisals and evaluations of outcomes using qualitative and quantitative data.

Economic appraisal and evaluation uses a variety of tools to estimate the costs and benefits of policy initiatives over time, such as the *discount rate* for adjusting the value of outcomes that will occur in the future. Detailed guidance on such tools, and on economic appraisal and evaluation more generally, are provided by *The Green Book* (published by HM Treasury) and in Chapter 8 of *The Magenta Book*.

Does Policy Evaluation Deal with Ethical Issues?

Policy making involves choices that are influenced by values and value judgements. Political decision making has to meet considerations about economic efficiency (i.e. the most effective use of scarce resources), and about social justice (e.g. the social distribution of the effects and outcomes of a policy, programme or project). Consequently, policy evaluation looks for structured and systematic ways of appraising the grounds upon which such decision making takes place.

Political philosophy and ethics provide such structured and systematic procedures for evaluating the values and value judgements that are at the heart of political decision making. Rawls' (1972) *Theory Justice* provides one approach involving what he calls the 'veil of ignorance'. This is a set of procedures for determining the social justice of political decision making. Nozick (1974) provides an alternative view about human rights and entitlements that has implications for how the state and the policy making process can, and should, function. Such considerations are often not considered by conventional texts on policy evaluation. *The Magenta Book* takes the view that guidance on policy evaluation should include some reference to the philosophical and ethical dimensions of policy making and policy implementation. Consequently, a discussion of how policy evaluation might incorporate such considerations is provided in Chapter 8.

How Does Policy Evaluation Relate to Project Management?

Policy evaluation and analysis requires a structured and organised approach to defining an answerable question, summoning appropriate and relevant evidence, critically appraising and analysing that evidence, identifying the risks and opportunities of a policy, programme or project, and determining the likely effects (positive and negative) of the project at hand. Project and programme management has emerged in recent years as a such structured and organised ways of planning, implementing and concluding projects and programmes. The congruity of interest between policy evaluation and project management is clear. Chapter 11 provides guidance on the main principles and procedures of project and programme management, and of their relevance to policy evaluation and analysis.

Summary

Policy evaluation is a family of research methods that are used to systematically investigate the effectiveness of policies, programmes, projects and other types of social intervention, with aim of achieving improvement in the social, economic and everyday conditions of people's lives. Different methods of policy evaluation are used to answer different questions. *The Magenta Book* provides a set of guidance notes on how to use the methods of policy evaluation and analysis effectively and, thereby, to generate and use sound evidence at the heart of policy making and implementation.

References

Chen, H.T. 1990

Theory-Driven Evaluations, Thousand Oaks, California, Sage Publications.

Cabinet Office (1999a) *Modernising Government*.

Cabinet Office (1999b) *Professional Policy Making for the 21st Century*

Cabinet Office (2000) *Adding it up: improving analysis and modelling in the 21st Century*

Cabinet Office (2001) *Better Policy Making*

Connell, J.P., Kubisch, A.C., Schorr, L.B. and Weiss, C.H., 1995

New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts, Washington D.C., Aspen Institute.

Funnell, S. 1997

'Program Logic: An Adaptable Tool', *Evaluation News and Comment*, **6**, 1, 5-17.

Greene, J.C., Benjamin, L., and Goodyear, L. 2001

'The Merits of Mixing Methods in Evaluation', *Evaluation*, **7**, 1, 25-44.

HM Treasury (2003) *Green Book: appraisal and evaluation in Central Government*

(http://www.hm-treasury.gov.uk/economic_data_and_tools/greenbook/)

Judge, J. And Bauld, L. 2001

'Strong theory, flexible methods: evaluating complex community-based initiatives', *Critical Public Health*, **11**, 1., 19-38.

Nozick, J. 1974

Anarchy, State and Utopia, Oxford, Basil Blackwell, 1974.

Office of Deputy Prime Minister/ Department for Transport (2002)

Integrated Policy Appraisal Framework

Owen, J.M. and Rodgers, P.J. 1999

Program Evaluation: Forms and Approaches, Thousand Oaks, California, Sage Publications.

Patton, M.Q. 1990

Qualitative Evaluation and Research Methods, 2nd ed, Newbury Park, California,

Sage Publications

Pawson, R. and Tilley, N. 1997

Realistic Evaluation, London, Sage.

Petrosino, A., Turpin-Petrosino, and Buehler, J., 2002

“‘Scared Straight’ and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency’, Campbell Library, www.campbellcollaboration.org.

Rawls, J. 1972

A Theory of Justice, Oxford, Clarendon Press, 1972.

Rogers. P.J., Petrosino, A., Huebner, T.A., and Hacsí, T.A. 2000

New Directions for Evaluation: Program Theory in Evaluation: Challenges and Opportunities, San Francisco, Jossey Bass Publishers.

Rossi, P.H., Freeman, H.E. and Lipsey, M.W., 1999

Evaluation: A systematic Approach, 6th Edition, Thousand Oaks, California, Sage Publications.

Scriven, M., 1991

Evaluation Thesaurus, 4th Edition, Newbury Park, California, Sage Publications

Weiss, C.H. 1997

‘Theory-based evaluation: past, present and future’, *New Directions for Evaluation*, 76, 41-55.

Filename: Chapter 1_final.doc
Directory: C:\WINDOWS\Desktop
Template: C:\Program Files\Microsoft Office\Templates\Normal.dot
Title: Chapter 1
Subject:
Author: nigel owens
Keywords:
Comments:
Creation Date: 08/07/03 13:58
Change Number: 13
Last Saved On: 08/07/03 03:05
Last Saved By: nigel owens
Total Editing Time: 36 Minutes
Last Printed On: 08/07/03 03:05
As of Last Complete Printing
Number of Pages: 13
Number of Words: 7,526 (approx.)
Number of Characters: 42,899 (approx.)