



AN INTRODUCTION TO EVALUATION

**CHRIS FOX
ROBERT GRIMM
RUTE CALDEIRA**



AN INTRODUCTION TO EVALUATION

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish over 900 journals, including those of more than 400 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne



AN INTRODUCTION TO EVALUATION

**CHRIS FOX
ROBERT GRIMM
RUTE CALDEIRA**

 **SAGE**

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne



Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Jai Seaman
Assistant editor: Alysha Owen
Production editor: Ian Antcliff
Copyeditor: Audrey Scriven
Marketing manager: Sally Ransom
Cover design: Shaun Mercier
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound by CPI Group (UK) Ltd,
Croydon, CR0 4YY

© Chris Fox, Robert Grimm and Rute Caldeira 2017

First published 2017

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Library of Congress Control Number: 2016935864

British Library Cataloguing in Publication data

A catalogue record for this book is available from
the British Library

ISBN 978-1-4739-0286-2
ISBN 978-1-4739-0287-9 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the PREPS grading system. We undertake an annual audit to monitor our sustainability.

TABLE OF CONTENTS

<i>About the authors</i>	vii
<i>Introduction</i>	ix
PART I: GETTING STARTED	1
1 What is evaluation?	3
2 The ethics of evaluation	23
PART II: UNDERTAKING AN EVALUATION	39
3 Theories of change	41
4 Process evaluation	61
5 Impact evaluation	77
6 Economic evaluation	113
7 Evaluation methods	139
PART III: THE PRACTICE OF EVALUATION	181
8 Planning an evaluation	183
9 Conducting an evaluation	205
PART IV: USING EVALUATION FINDINGS	221
10 Systematic reviews	223
11 Knowledge mobilisation: getting evidence into policy and practice	241
PART V: EVALUATION PARADIGMS	259
12 Evaluation paradigms and the limits of evidence-based policy	261
13 Conclusion	279
<i>Glossary</i>	283
<i>References</i>	291
<i>Index</i>	307

ABOUT THE AUTHORS

Chris Fox is Professor of Evaluation and Policy Analysis at Manchester Metropolitan University, where he is Director of the Policy Evaluation and Research Unit (www.mmu-peru.co.uk). PERU is a multi-disciplinary team of evaluators, researchers, analysts, and economists and undertakes evaluation and applied research in the UK and Europe. Chris's work cuts across a number of policy areas. He has led numerous evaluations in the criminal justice system and currently is leading large, long-term research and evaluation programmes on reducing re-offending, and on prisoner education. Chris is also interested in social policy innovation and is Director of a European Commission-funded project 'Innovative Social Investment: Strengthening communities in Europe' that is studying innovative approaches to welfare reform in ten European countries.

Robert Grimm is Associate Director at Ipsos and leads the Ipsos Public Affairs, Political and Social Research Team in Germany. In this capacity, Robert is responsible for the design and management of major social research projects for academic institutions, NGOs and Private Sector Clients. His work includes developing research instruments, sampling strategies, designing indicators and measurements, setting up baselines measures and continuous performance tracking. Robert is also leading on the development of Ipsos' political polling in Germany. Before joining Ipsos, Robert worked as an academic at Manchester Metropolitan University. Robert's research interests included social innovation and social investment and he has been involved in many European Commission funded research projects. Robert taught evaluation and social research methods to post-graduate students and continues to be a fellow at Manchester Metropolitan University.

Rute Caldeira is a social scientist with a doctorate in sociology. She has 15 years of work experience, mostly in the NGO sector; the last eight years were dedicated to setting up planning, performance and monitoring systems both in complex not-for-profit and in community-based organisations with a view to enhance their ability to enable social impact. Rute currently heads Transparency International Monitoring, Evaluation and Learning Unit, which she set up in 2011. She brings in extensive knowledge of methods and methodological approaches, of evaluations and assessments, and of organisational capacity to absorb learning from impact studies and evaluations.

INTRODUCTION

WHO SHOULD READ THIS BOOK?

This book has been written with two main audiences in mind.

The first group is students, in particular postgraduates doing taught courses in research methods and both undergraduates and postgraduates planning research projects, perhaps for a final year dissertation or a PhD.

The second group is people working in governmental or non-governmental organisations who are planning either to undertake an evaluation or commission an evaluation.

WHY EVALUATION IS IMPORTANT

Evaluation is a fascinating interdisciplinary research practice with a wide spectrum of applications. **Evaluators** continuously encounter new challenges; they work in changing environments, with a variety of people and organisations where they have to think themselves into new social settings. No single evaluation is like another and evaluators need to pragmatically develop innovative and robust research methods.

At a time of increasing calls for transparency and accountability, evaluation has moved centre stage in public policy, organisational planning and management. Government departments are under pressure to deliver social policy programmes that are effective *and* value for money. Chelimsky (2006) suggests that evaluation is an intrinsic part of democratic government for four reasons:

It reports information about government performance that the public needs to know. It adds new data to the existing stock of knowledge required for government action. It develops an analytical capability within agencies that moves them away from territoriality and toward a culture of learning. And, more generally, its spirit of scepticism and willingness to embrace dissent help to keep government honest. (Chelimsky 2006: 33)

Similarly, non-governmental organisations from Amnesty International to the Red Cross are accountable towards their funders and have to prove that they use donations in an appropriate way. Private sector organisations too are increasingly evaluating internal processes and operations. The evaluator has an important role in assessing

and judging programmes, policies and interventions, and evaluation reports often determine the future of interventions.

Evaluation combines managing the expectations of different **stakeholders** – those who are affected by an intervention, those who help to deliver the intervention and those who have been funding it – with understanding the social complexities and remaining objective and neutral. While handling various stakeholders is challenging, understanding diverse points of views is also insightful and rewarding. Evaluation research requires multiple skill sets from project planning, team building and stakeholder management to the design of interdisciplinary and multi-method research instruments. It is this diversity and the applied nature of evaluation research which make it an exciting and stimulating task.

Good evaluators are not just expert social scientists, they are also expert negotiators and project managers. They have a deep knowledge of social science method and theory, but are also practical and politically astute.

OUR APPROACH TO EVALUATION

A practical discipline

Methodological rigour, theory development and a coherent philosophy are all important in evaluation, but above all it remains a practical discipline. Academic debate about the merits of different **evaluation paradigms**, types of data collection or approaches to analysis must be tempered by the needs of different stakeholders, including funders, policymakers and service users. Unlike some other types of social science research, evaluators almost always work within practical constraints such as fixed budgets and deadlines.

Evidence-based (or informed) policy and practice

One of the distinguishing features of evaluation is that it is undertaken to improve programme effectiveness and/or inform decisions about future programmes (Patton 1997). However, this raises lots of challenges and questions for evaluation, including the competing needs and interests of various stakeholders and the extent to which evidence can or should influence policy and/or practice. Effective planning, good project management and a proper strategy for knowledge mobilisation are just as important as a well-designed **evaluation framework** and the appropriate use of research methods.

Multi-sector, multi-disciplinary

Evaluation has an established or growing profile across a wide range of sectors including education, welfare, health and social care, criminal justice and international

development. Different approaches to evaluation are prominent in different sectors. In writing this book we have tried to draw on examples from a range of sectors in order to illustrate the range and diversity of **evaluation theory** and practice and encourage cross-pollination of ideas.

Complexity

Possibly the biggest challenge facing evaluators today is how to respond to complexity. We think of complexity in two senses.

Firstly, across a number of sectors there is increasing recognition that service users often have multiple and ‘complex’ needs and that services often have to be organised that involve collaboration between various organisations. Here we are recognising that people and programmes are *complicated*. So, for example, a housing need may be linked to a substance misuse need that is, in turn, linked to a mental health need. Alternatively, an individual child’s educational needs may be interlinked with a broader set of health and social care needs within their family and those needs may be intergenerational. These needs require multiple delivery partners working across different sectors and service delivery often takes place within a fast-moving policy environment where change is the norm.

Secondly, we also think of complexity in a sociological sense. This goes beyond the idea that programmes are *complicated* and embraces the concept of *complexity* (Glouberman and Zimmerman 2002). Complexity includes the principle of non-linearity (small changes in inputs may, under some conditions but not others, produce large changes in outcome); the contribution of local ‘adaptiveness’ and feedback loops; the phenomenon of emergence; the importance of path dependence; and the role of human agency (Marchal et al. 2013).

In both of these senses, complexity poses challenges for evaluators seeking to explain how interventions work, what impacts they have, and whether they make economic sense.

WHAT DOES THIS BOOK COVER?

This book includes:

- practical steps for designing and conducting an evaluation and ensuring that the evaluation is ethical
- different approaches to evaluation, including theories of change, **process evaluation, impact evaluation (outcome evaluation)** and **economic evaluation**
- ways of reviewing evaluation evidence including systematic reviews

- how to get evaluation knowledge into policy and practice
- important philosophical questions about evaluation and its role in society

The book is divided into several sections.

Part I: Getting Started

This section introduces evaluation, providing a brief overview of the field and highlighting some key issues to address before starting an evaluation.

Chapter 1 introduces the field of evaluation and provides an overview of different types of evaluation and the historical development of the discipline.

Chapter 2 considers the **ethics** of evaluation, both the design of ethical evaluations and ethical evaluation practice.

Part II: Undertaking an Evaluation

This section covers common components of an evaluation. Not every evaluation will use all these components but most evaluations will use some of them.

Chapter 3 introduces Theories of Change, an important tool for evaluators and programme developers and an approach that can be integrated into many evaluations.

Chapter 4 discusses **process (implementation) evaluation**.

Chapter 5 discusses impact (outcome) evaluation.

Chapter 6 discusses economic evaluation.

Chapter 7 introduces a range of quantitative and qualitative methods for data collection and analysis commonly used in evaluation. **Mixed methods** (dual strategies) are also considered.

Part III: The Practice of Evaluation

This section provides practical guidance on the practice of evaluation.

Chapter 8 gives practical guidance and advice on planning an evaluation, with a particular focus on refining evaluation questions and engaging with different stakeholder groups.

Chapter 9 provides practical guidance and advice on conducting an evaluation, with a particular emphasis on the importance of good project management.

Part IV: Using Evaluation Findings

Over recent years, increasing emphasis has been placed on the use of evaluation findings.

Chapter 10 describes systematic reviews, which are used increasingly to synthesise evaluation findings for policymakers and practitioners.

Chapter 11 gives practice guidance on disseminating evaluation results and also discusses how to include evaluation evidence in policy and practice (knowledge mobilisation).

Part V: Evaluation Paradigms

Chapter 12 provides an introduction to some of the philosophical debates that underpin evaluation, including a discussion of three commonly used evaluation paradigms.

Chapter 13 offers some concluding thoughts on the theory and practice of evaluation.

HOW TO USE THIS BOOK

The book can either be read from start to finish or if you have particular interests it is possible to concentrate on particular chapters:

- For a quick overview of evaluation, go to Chapter 1
- If you are designing an evaluation either as an evaluator or as a commissioner, concentrate on Chapters 3–7
- For practical advice on planning an evaluation and then conducting an evaluation, look at Chapters 8–9
- If you are thinking about reviewing existing evaluation findings or using evaluation evidence to influence policy and practice, look at Chapters 10–11
- If you want a deeper understanding of different evaluation paradigms and some of the key philosophical debates within evaluation, look at Chapter 12

PART I

GETTING STARTED

1

Introduction	4
Defining evaluation	4
Different types of evaluation	9
Trends in evaluation	15

WHAT IS EVALUATION?

WHAT IS EVALUATION?

INTRODUCTION

Evaluation is a broad concept and one that is sometimes difficult to distinguish both from other types of research and from related practices such as **monitoring**, performance management and audit. We start this chapter by looking at the various definitions of evaluation, distinguishing those that concentrate on the purpose of evaluation, the methods used in evaluation, and the significance of judgements of value in evaluation. We consider each of these briefly before adopting a definition of evaluation based on judgements of value. This is a definition that emphasises the political dimension of evaluation – a key distinguishing feature that will help us differentiate it from other types of research activity.

Evaluation covers a range of activity types. People commonly talk and write about formative and **summative evaluation**, process (implementation) and impact (outcome) evaluation, and economic evaluation and theory-led evaluation. All of these terms raise questions about the nature and role of evaluation which we will address at various points in the book. In this chapter we will look at types and models of evaluation before discussing some key trends that have shaped the world of evaluation over recent decades.

The final section of this chapter looks briefly at the history of evaluation. It aims to introduce readers new to evaluation to some of the ‘big’ ideas that have helped shape the sector and are still debated today. In particular the flux between ‘scientific’ and ‘naturalistic’ approaches to evaluation is discussed. Many of the ideas raised in this chapter will be developed in detail in later chapters.

DEFINING EVALUATION

Mark et al. (2006) distinguish everyday informal evaluation (How good was breakfast at the restaurant? How did the meeting with the client go?) from systematic evaluation, which they define as:

a social and politicized practice that nonetheless aspires to some position of impartiality or fairness, so that evaluation can contribute meaningfully to the well-being of people in that specific context and beyond. (Mark et al. 2006: 5–6)

They identify three groups of evaluation definitions which concentrate on the purpose of evaluation, the methods used, and the importance of judgements of value in evaluation. We will consider each of these in turn.

Defining evaluation according to purpose

Mark et al. (2006) identify a group of definitions that concentrate on the purpose of evaluation, typically providing information for policymaking or programme improvement. Their example is Patton’s definition:

Program evaluation is the systematic collection of information about the activities, characteristics, and **outcomes** of programs to make judgments about the program, *improve program effectiveness, and/or inform decisions about future programming*. (Patton 1997: 23, emphasis added)

A note on terminology: policy, programme or project?

At this point it is useful to note that in the literature on evaluation we often come across references to projects, programmes and policies. Building on Eggers' work, Palfrey et al. (2012) argue that it is necessary to distinguish between these three evaluation subjects because they might offer the evaluator different opportunities to contribute to decision making. They suggest that:

- a project is a planned activity aimed at achieving specified goals within a prescribed period
- a programme is a set of separate planned activities unified into a coherent group
- a policy is a statement of how an organisation or government would respond to particular eventualities or situations according to its agreed values of principles

In this book, while we accept that projects, programmes and policies present different evaluation opportunities and challenges, for the sake of brevity our default position will be to refer to 'programme' evaluation unless there is a need to distinguish one from another.

Defining evaluation according to method

Another group of evaluation definitions identified by Mark et al. (2006) outline evaluation in terms of methods. An example comes from Rossi et al.:

Program evaluation is *the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions*. (Rossi et al. 2004: 16, emphasis added)

Defining evaluation in terms of methods can be potentially helpful in distinguishing it from similar practices such as monitoring, performance management, auditing and accreditation. However, in turn this raises questions about what distinguishes evaluation from research (Palfrey et al. 2012).

Evaluation distinguished from monitoring, performance management, audit and accreditation

Monitoring, performance indicators (PIs) and broader performance management process have proliferated across the public, private and, increasingly, the not-for-profit

sectors in the last few decades. We can link this proliferation to the development of 'New Public Management'. Often associated with reforms to the public sector that were introduced during the administrations of Prime Minister Margaret Thatcher in the UK and President Ronald Reagan in the US, New Public Management (NPM) involved structural changes to the public sector and the introduction of business methods into government (Hill and Hupe 2014), as well as practices such as shrinking the size of the state so that government reduced its service delivery capacity, contracting out government services to the private and not-for-profit sectors, a greater emphasis on the choice available to service users, and the creation of new public-private vehicles to deliver services (Hill and Hupe 2014). The adoption of business practices, a greater focus on managing by **outputs** and the increased 'marketisation' associated with NPM contributed to the proliferation of performance management measures.

Palfrey et al. (2012: 19, citing Carter et al. 1992) suggest that PIs are very useful as 'tin openers' because they help us clarify questions about performance. In this sense, they are a valuable starting point for evaluation, but are not a substitute for evaluation that incorporates the use of research methods (Palfrey et al. 2012).

Audit has also proliferated in the UK and US. Power (1997) charts a move from traditional audits that focus on financial probity to audits that ask broader questions about organisational performance and 'Value for Money' (VFM) (Palfrey et al. 2012). However, deciding on what matters in VFM involves value judgements (ibid.) and by imposing values audits can have unintended and dysfunctional consequences for the audited organisation. Evaluation does not avoid such value judgements but social scientists recognise their importance and have developed a number of strategies to avoid or incorporate them, depending upon the social science tradition they come from.

Accreditation has been used widely in the UK public sector as a strategy for setting standards for the performance of organisations and often starts with self-evaluation (Palfrey et al. 2012). Well-known examples in the UK include the use of 'Trust' status in sectors such as health and 'Investors in People' – a government agency that accredits organisations that demonstrate good practice in workforce management (ibid.).

If the 'research' component is what distinguishes evaluation from practices such as PIs, audits and accreditation, what is it then that differentiates 'evaluation' from 'research'?

Distinguishing evaluation from research

The distinction between evaluation and research is discussed by Lincoln and Guba (1986) who note that both are forms of 'disciplined inquiry' and use many of the same tools or methods. However, having such shared methods does not make

them one and the same thing. Lincoln and Guba argue that ‘to assert identity or similarity on the basis of common methods would be analogous to saying that carpenters, electricians, and plumbers do the same thing because their tool kits all contain hammers, saws, wrenches, and screwdrivers’ (1986: 547). The key distinction between evaluation and other types of research is the importance of values of judgement in evaluation.

Defining evaluation according to judgements of value

This brings us to the third group of definitions identified by Mark et al. (2006), which concentrate on the function evaluation serves and assume that evaluation involves judgements of value. As an example of a definition of evaluation based on judgements of value, Mark et al. (2006) cite Scriven’s definition:

Evaluation refers to the process of *determining the merit, worth, or value of something*, or the product of that process ... The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of the **evaluands** on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations. (Scriven 1991: 139; emphasis added)

Lincoln and Guba (1986), when considering what makes evaluation different from research, argue that latter is undertaken to resolve a problem while evaluation is undertaken to establish value. They also define research as: ‘a type of disciplined inquiry undertaken to resolve some problem in order to achieve understanding or to facilitate action’ (1986: 549), whereas evaluation is defined as:

a type of disciplined inquiry undertaken to determine the value (merit and/ or worth) of some entity – the evaluand – such as a treatment, program, facility, performance, and the line – in order to improve or refine the evaluand (**formative evaluation**) or to assess its impact (summative evaluation). (1986: 550)

This difference, which Lincoln and Guba describe as ‘monumental’, also leads to what they see as a key distinction in the products that result. Whereas research is typically adequately served by a technical report, this by itself is rarely sufficient for an evaluation if it has to meet the needs of, and communicate with, its various audiences (Lincoln and Guba 1986).¹

¹Arguably, this distinction in products is diminishing as researchers and research funders place more emphasis on research achieving ‘impact’ and research knowledge being ‘mobilised’ more effectively.

Our preferred definition of evaluation

The many definitions of evaluation suggest that it is not easy to pin down the concept. Indeed, some observers have argued that this is a pointless task. For example, some evaluators reject objective 'scientific' approaches to evaluation, arguing instead that because the human world is socially constructed evaluation is itself a social construct. There are multiple social constructs and therefore from this relativist point of view there is no right way to define evaluation. Thus, by the end of the decade, Guba and Lincoln were arguing: 'There is no answer to the question, "But what is evaluation really?" and there is no point in asking it' (1989: 21).

We recognise the importance of purpose and methods in defining evaluation, but also take the view that what is crucial for defining evaluation is the emphasis on a process of determining the merit, worth or value of something, along the lines suggested by Scriven.

Distinguishing evaluation from research as a practice designed to establish the value of an entity has notable implications that will resurface throughout this book. If we accept that the purpose of evaluation is to determine the value of the entity being evaluated, and that the products of an evaluation are designed to improve the thing being evaluated or to assess its impact, this has important repercussions for evaluation and for evaluators. If we return to the very first definition of evaluation that we considered, i.e. Mark et al.'s (2006) view of evaluation as a 'politicized practice that nonetheless aspires to some position of impartiality or fairness', we can start to see the potential tensions in a practice that is at once politicised but also aspires to impartiality or fairness.

Some would go further and see in the literature the suggestion that evaluation is either a political activity or that it serves a political purpose (Palfrey et al. 2012). In their review of the relevant literature Palfrey and colleagues distinguish between these two possibilities. Citing the work of Patton (1988) they suggest that, at a minimum, if evaluation is intended to improve services then, in the public sector, the decision makers who will act on evaluation findings are either local or national politicians. In this case evaluation serves a political purpose. However, many commentators on evaluation go further and, as Palfrey et al. (2012) note, with particular reference to the collected work of Guba and Lincoln, some see evaluators as ideologically committed with political sympathies that *should* influence the design of their evaluations. Whichever view is taken, they argue that over recent decades:

evaluation has come to be associated for the past few decades as a potent ally of politicians in that it is a means of assessing the 'value' of projects, programmes and policies. (Palfrey et al. 2012: 29)

The implication of this is that:

In the real world of politics, despite the mass of literature supporting and promoting evaluation as a subject worthy of intense bookish activity, it has to be acknowledged that for all its intellectualising credentials it is a servant and not an equal of politicians. (2012: 29)

DIFFERENT TYPES OF EVALUATION

A number of distinctions are made when describing different types of evaluation and we look at some of the more common ones here.

Formative and summative evaluation

Scriven (1967) makes a distinction between formative and summative evaluation, which Lincoln and Guba (1986) suggest are, broadly speaking, aims of evaluation:

The aim of formative evaluation is to provide descriptive and judgmental information, leading to refinement, improvement, alterations, and/or modification in the evaluand, while the aim of summative evaluation is to determine its impacts, outcomes, or results. (Lincoln and Guba 1986: 550)

A less technical, but similar definition is provided by Robson:

Formative evaluation is intended to help in the development of the programme, innovation or whatever is the focus of the evaluation. Summative evaluation concentrates on assessing the effects and effectiveness of the programme. (Robson 2011: 181)

However, the distinction between summative and formative evaluations is not absolute (Robson 2011). For example, determining whether or not a policy has had an impact often involves asking questions about how it has done so, for whom, why, and under what conditions (Government Social Research Unit 2007a). These two broad approaches tend to carry with them some assumptions about the nature of the evaluation undertaken. For example, a formative evaluation needs to be carried out and reported in time for modifications to the policy, programme or project (Robson 2011), implying that the audience for the evaluation might more likely be programme managers as opposed to policymakers or the public. This in turn implies the evaluator might have a more interactive role and that data collection will be continuous, possibly with a greater emphasis on qualitative data collection. A summative evaluation implies that the audience are more likely to be policymakers, programme funders or the public, and that the role of the evaluator could be more independent and removed from evaluation subjects, with a focus on outcome

measures and the production of formal evaluation reports. Potential distinctions are illustrated in Table 1.1.

Table 1.1 Distinguishing a formative and summative evaluation (based on Scriven 1967 and Herman et al. 1987)

	Formative	Summative
Target audience	Programme managers/practitioners	Policymakers/funders the public
Focus of data collection	Clarification of goals Nature of implementation	Identifying outcomes
Role of evaluator	More likely to be interactive and involved	More likely to be independent or 'arm's length'
Methodology	Quantitative and qualitative (emphasis on latter)	Emphasis on quantitative
Frequency of data collection	Continuous monitoring	At intervals
Reporting procedures	Both formal and informal via discussion groups and meetings	Formal reports
Frequency of reporting	Throughout period of observation/study	On completion of evaluation

Process (implementation) and impact (outcome) evaluation

Another common distinction in the evaluation world is that between process (sometimes referred to as implementation) and impact (sometimes referred to as outcome) evaluation. Traditionally, evaluation was restricted to questions concerning impact or outcome and typical among these might be (based on HM Treasury 2013):

- What were the policy, programme or project outcomes?
- Did the policy, programme or project achieve its stated objectives?
- Were there any observed changes, and if so how big were these changes and how much could be said to have been caused by the policy, programme or project as opposed to other factors?
- How did any changes vary across different individuals, stakeholders, sections of society and so on, and how did these compare with what was anticipated?
- Did any outcomes occur which were not originally intended, and if so, what were they and how significant were these?

Process evaluation answers the question 'How was the policy, programme or project delivered?' (HM Treasury 2013) or the 'What is going on?' question (Robson 2011). Process evaluation may provide a useful additional element to an outcome evaluation, helping to explain *why* an intervention did or didn't work. However, a process

evaluation may be a ‘standalone’ exercise designed to provide a detailed description of the implementation process or a counterpoint to an ‘official’ view of what should be happening in a policy, programme or project (Robson 2011).

Impact evaluation therefore looks similar to summative evaluation and process evaluation looks similar to formative evaluation, but there are distinctions. For example, a summative evaluation occurs at the end of a programme, whereas an impact evaluation need not necessarily do so.

A note on terminology: impact versus outcome and process versus implementation

In this book we generally use the term ‘impact evaluation’ rather than ‘outcome evaluation’. The two terms overlap, but have slightly different meanings. An outcome evaluation assesses whether a programme delivered outcomes specified in the programme, whereas an impact evaluation is broader and also considers unintended or wider outcomes. We also generally use the term ‘process evaluation’ rather than ‘implementation evaluation’. Again, the two terms overlap, but whereas an implementation evaluation assesses whether a programme was delivered as intended, a process evaluation also considers unintended or wider delivery issues.

Economic evaluation

A summative or outcome evaluation might demonstrate the impact of a policy, programme or project but will not by itself show whether those outcomes justified the investment (HM Treasury 2013). Evaluators may ask (based on Dhiri and Brand 1999):

- What was the true cost of an intervention?
- Did the outcome(s) achieved justify the investment of resources?
- Was this the most efficient way of realising the desired outcome(s) or could the same outcome(s) have been achieved at a lower cost through an alternate course of action?
- How should additional resources be spent?

In general, attempts to address these issues fall into one of three forms:

- Cost Analysis is a partial form of economic evaluation that deals only with the costs of an intervention (Drummond et al. 2005).
- **Cost Effectiveness Analysis** values the costs of implementation and relates these to the total quantity of outcome generated to produce a ‘cost per unit of outcome’ estimate. The consequences of the policy, programme or project are not valued and the results are expressed as a cost-effectiveness ratio such as ‘the cost per additional individual placed in employment’ (HM Treasury 2013).

- **Cost Benefit Analysis** (CBA) goes further than a cost effectiveness analysis and the consequences of the policy, programme or project are valued in monetary terms (Drummond et al. 2005). Results are often expressed as a cost-benefit ratio such as ‘for every dollar spent placing an individual in employment there is a return of two dollars’. Potentially this makes it the broadest form of economic evaluation method, however, as we will discuss later, difficulties in capturing and measuring the wider consequences of an intervention means that, in reality, its scope can be limited.

We look at economic evaluation in more detail in Chapter 6.

Ex ante and ex post evaluations

Evaluation can be prospective or retrospective. A *prospective* or **ex ante evaluation** takes place before a programme or project has been implemented, whereas a *retrospective* or **ex post evaluation** takes place once a programme or project is in place and has demonstrated that it has had an impact (Rossi et al. 2004).

Ex ante evaluations are most commonly undertaken by governments or similar bodies as part of the policy and programme development cycle. They normally have a strong economic component. The European Commission (2001) defines ex ante evaluation thus:

Ex ante evaluation is a process that supports the preparation of proposals for new or renewed Community actions. Its purpose is to gather information and carry out analyses that help to define objectives, to ensure that these objectives can be met, that the instruments used are cost-effective and that reliable later evaluation will be possible. (European Commission 2001: 3)

Ex post evaluation is far more common than ex ante evaluation and the bulk of this book concentrates on approaches more commonly associated with ex post evaluation.

The distinction between ex ante and ex post evaluation alerts us to the idea that different types of evaluation will be relevant at different points in the policymaking or programme development process. The Public Service Transformation Network (2014) identified a series of stages in a project or programme life-cycle: development and design; implementation; delivery; and scaling-up. They suggest that various types of evaluation are likely to be relevant at different stages in this lifecycle, as illustrated in Figure 1.1.

However, the underlying assumption in this model is that policymaking is a rational process. The rational model suggests a logical and ordered sequence of policymaking phases that follow a linear process and where evaluation knowledge is apolitical and neutral (Stone et al. 2001). However, many would argue that the reality of policymaking is more complicated. Clearly, policymakers do not just rely on evidence generated from robust evaluations: they bring their own experience, expertise and judgement

WHAT IS EVALUATION?

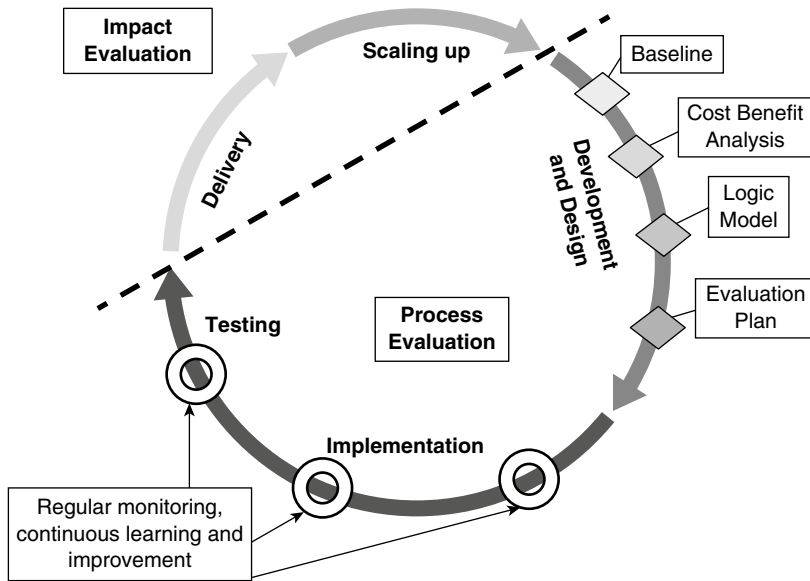


Figure 1.1 Evaluation through the lifecycle of a policy or programme (source: Public Service Transformation Network 2014, Figure 1.1)

to the process (Davies 2004). Policymaking also takes place within the context of finite resources and this influences decisions (ibid). Policymakers are also influenced by values (their own, as well as those of politicians and of institutions) along with the habits and traditions of institutions such as Parliament, civil servants and the judiciary (ibid). Outside forces such as lobby groups, pressure groups and consultants are able to influence the policymaking process and that whole process is subject to pragmatics and contingencies such as parliamentary terms and timetables and the capacities of institutions (ibid). Thus many commentators prefer the phrase ‘evidence-informed policy’ (e.g. Treadwell Shine and Bartley 2011).

However, some would go further and question the extent to which **evidence-based policymaking** actually takes place at all. Stone et al. (2001) set out alternative models of policymaking where the role of evidence and of evaluators is either marginal (for example, in a ‘muddling through’ model of policymaking premised on the notions of ‘bounded rationality’ and ‘satisficing’) or only influential at certain stages in the development of a ‘policy paradigm’.

Theory-led evaluation

Donaldson and Lipsey note that:

Reference to theory is widespread in the contemporary evaluation literature, but what is meant by ‘theory’ encompasses a confusing mix of concepts ... A newcomer

to evaluation, and even a grizzled veteran, could have a difficult time sorting through the closely related and sometimes interchangeable terms that litter the evaluation landscape ... (Donaldson and Lipsey 2006: 57)

As Donaldson and Lipsey note, the challenge here is partly that terms are used inconsistently and with overlapping meanings, but also partly that the role of theory in evaluation is contentious. They suggest that there are three common types of theory encountered in evaluation:

- Evaluation theories are theories of evaluation practice that address questions such as how to understand the nature of what we evaluate, how to assign value to the things we evaluate and their performance, how to construct knowledge, and how to use the knowledge generated by evaluation.
- Social science theories attempt to explain the social world, and when they address social phenomena related to social programmes and the social conditions they are intended to improve, those social theories can be very relevant to evaluation.
- Theory-based evaluation places emphasis on developing **programme theory** as an integral component of the evaluation process. It is more modest than social theory and deals with the assumptions that guide the way specific policies, programmes and projects are implemented, and are expected to bring about change.

Evaluation theory

Evaluation theory is largely focused inwards on the evaluation profession and helps evaluators discuss their practice with each other (Donaldson and Lipsey 2006, Shadish 1998). However, evaluation theories can also help evaluators explain various evaluation approaches and practices to those commissioning evaluations, thereby helping find an optimal fit between an **evaluation team** and the needs and interests of the evaluation commissioner (Donaldson and Lipsey 2006). Alkin and Christie (2004), building on work by Alkin (2004), developed an ‘evaluation theory tree’. They looked at the work of a number of prominent evaluators and placed each of these on one of three branches of a tree designed to indicate the three streams or traditions in evaluation theory. One concentrated on evaluation methods, another on how data were to be judged or valued, and the third on users and use (Mark et al. 2006). The branches represented the evaluator’s *relative degree of emphasis* (ibid.) on the three issues, and in the graphical representation some evaluators sit close to the junction of two branches while others are located at the far end of a branch. Mark and colleagues suggest the evaluation theory tree is useful in highlighting some of the major conceptual emphases in the field of evaluation. We consider evaluation theory in more detail in Chapter 12.

Social science theory

Donaldson and Lipsey (2006) see several important roles for social science theory in evaluation including informing initial **needs assessment** and programme design, helping evaluators assess the likelihood that programmes will be able to accomplish certain objectives, and guiding evaluation measurement and design issues. Examples of social science theories given by Donaldson and Lipsey include social cognitive learning theory, the theory of planned behaviour, the theory of health behaviour change, and theories of learned helplessness suggesting that ‘mid-level theories’ are likely to be of most relevance to evaluators.

Theory-based evaluation

Donaldson and Lipsey (2006) suggest that a well-developed programme theory is useful for framing key evaluation questions and designing sensitive and responsive evaluations. This process often involves programme stakeholders in the articulation of programme theory. While most evaluators would agree that a well-developed programme theory is important some groups of evaluators would go further. In recent years both theories of change and **realistic evaluation** have developed rapidly and they possess some similarities: both see limitations in methods-driven approaches to evaluation design, see theory development as critical to the evaluation process, and emphasise the importance of programme context in understanding how programmes lead to changes in outcomes (Blamey and Mackenzie 2007). And yet they are not entirely interchangeable approaches, use theory in slightly different ways, and arguably are differently suited to evaluating projects, programmes and policies (ibid.). In Chapter 3 we look in more detail at programme theory and in particular at a popular evaluation tool – ‘theories of change’.

TRENDS IN EVALUATION

Mark et al. (2006), in their brief commentary on the history of evaluation, argue that there is not *a single history* of evaluation but instead *multiple histories* of evaluation depending upon one’s discipline and domain of work. Below, we look at trends in evaluation through a number of different lenses.

Evaluation ‘booms’

Evaluation has a relatively short history when compared to that of the wider social sciences. Some histories of evaluation, particularly with a US focus, point to the

evaluations of social programmes in sectors such as education and public health prior to the First World War as the start of the modern history of evaluation, characterised by the application of social research methods to programme evaluation (see, for instance, Rossi et al. 2004). By the 1930s, social scientists were using rigorous methods to assess a range of social programmes, with the Second World War providing a strong boost as evaluators worked with the US Army to develop procedures for monitoring soldier morale and evaluating personnel policies and propaganda techniques (ibid.).

The first 'boom' in evaluation

Donaldson and Lipsey (2006) note that the first major boom in evaluation seemed to occur in the United States in the late 1960s and 70s under the Kennedy and Johnson administrations, when large social programmes funded by Federal government led to the development of experimental and **quasi-experimental** approaches to evaluation. These are evaluation designs that involve the use of **control groups**, and in the case of **experiments** a random allocation of evaluation subjects between an intervention and a control group. They are designs that draw on the tradition of laboratory experiments in disciplines such as medicine, psychology and agriculture. We discuss these evaluation designs in detail in Chapter 5. This first, post-war, boom period was also characterised by developments in social research methods, including sample surveys and advanced statistical procedures (Rossi et al. 2004). An evaluation community was emerging by the 1970s, and the first journal in evaluation, *Evaluation Review*, was launched in 1976 by Sage Publications (ibid.).

The second 'boom' in evaluation

Building on their previous work, Donaldson and Lipsey (2006) describe a 'second boom in evaluation' which happened during the last years of the twentieth century and the early years of the twenty first. This is more global and includes a wider range of governmental and non-governmental organisations commissioning evaluations and a growing number of evaluation professionals and professional associations. Evaluation as a practice has become globalised principally due to the link between effectiveness, transparency and accountability. Another characteristic of this second boom is the development of new theories of evaluation, new evaluation methods and new evaluation tools designed to address a broader and more diverse range of evaluation practice challenges (Donaldson and Lipsey 2006). This also represents a shift with evaluation, no longer shaped mainly by the interests of evaluators and consumers of evaluation exercising significant influence (Rossi et al. 2004). For Rossi and colleagues

the incorporation of the consumer perspective has moved evaluation beyond academic social science:

Evaluation has now become a political and managerial activity that makes significant input into the complex mosaic from which emerge policy decisions and resources for starting, enlarging, changing or sustaining programs to better the human condition. (Rossi et al. 2004: 10)

An implication of this move is that evaluation must be seen as an integral part of the social policy and public administration.

Real time evaluation

Real time evaluation (RTE) as a concept and practice took greater hold in the humanitarian sector. This was driven by a need to produce rapid assessments of interventions in emergency situations, where time is of the essence and failures are not opportunities for learning. Rather, programme failures in emergency contexts may lead to dramatic consequences. The focus is mostly on process. Additionally:

Unlike the majority of final ex-post evaluations, the process and products of an RTE are integrated within the programme cycle. Interaction with programme staff and managers during the course of implementation means that discussion, which may or may not be reflected in a final document, can help to bring about changes in the programme, rather than just reflecting on its quality after the event. (Herson and Mitchell 2005: 43)

More recently, Oxfam began trialling real time evaluations in its interventions in fragile and conflict-affected states and in crisis situations, i.e. in rapidly changing contexts that were unstructured and unpredictable, much as in emergency situations.

RTE embodies the rapid metamorphosis that evaluations go through notably in certain sectors like the humanitarian and international development ones. The latter are increasingly under pressure to demonstrate effectiveness and impact due to either the squeeze in countries' Official Development Assistance (ODA) or change in the prioritisation of countries' ODA allocations. However demonstrating effectiveness and impact in complex, rapidly changing contexts, where the availability of and access to data can be a challenge, has led to a proliferation of evaluation approaches that do not strictly conform to textbook instructions.

Evaluation, social policy and public administration

Several commentators have charted the relationship between the evaluation sector and changing fashions in government.

The growth of government

Rossi et al. (2004) note that the expansion of evaluation accompanied the (US) expansion of government, particularly at the federal level. They note that in the 1930s the Great Depression in the US saw rapid growth in ‘human services’ provided by government, and during that same period social scientists started to investigate the political, organisational and administrative decision making that took place in government. As government became increasingly complex and technical the importance of evaluation started to be acknowledged by politicians and administrators (ibid.). As described above, evaluation continued to expand rapidly during the Kennedy and Johnson administrations in the 1960s in response to large-scale social programmes such as the War on Poverty and the Great Society (ibid.). This was growth primarily in summative, goals-based evaluation – what Vedung (2010) describes as the ‘science-driven wave’:

From its inception in the late 1950s and consolidation in the mid-1960s, evaluation has been embedded into one of the great narratives of our time: that the world can be made more humane if capitalism and the market economy can be reined in by appropriate doses of central policy planning and public intervention at a comprehensive level ... In public-sector thinking, this was hailed as a victory of a kind of rationality. Public policy should be made more scientific and sensible. (Vedung 2010: 265)

Neo-liberalism and new public management

However, in the post-war period the dominance of the *scientific* approach to evaluation most clearly expressed in experimental evaluation designs was challenged on a number of fronts. In some branches of the social sciences the positivist paradigm that the scientific approach rests upon was questioned (we discuss this in more detail in Chapter 12) and throughout the 1970s and 80s evaluation theory saw the increasing influence of more pluralistic and naturalistic approaches to evaluation (Lincoln and Guba 1986). In the US, the Reagan administration of 1980 signalled the start of attempts to curtail domestic federal expenditure (Rossi et al. 2004), and in the UK Thatcher’s government of 1979 introduced a new approach to public sector accountability that since has become known as ‘the new public management’ or ‘managerialism’ (Palfrey et al. 2012), a trend that continued through successive governments including the Blair administration of 1997 (Peck 2004), the Brown administration and the Conservative government under the stewardship of David Cameron. This new emphasis on accountability and ‘value for money’, or what Vedung refers to as the ‘neo-liberal wave’, has, if anything, strengthened the position of evaluation, although at the same time it has changed it:

In the neo-liberal wave, it is regarded as imperative that the fundamental principal in a representative democracy, the demos, has a right to know how her agents spend her