2nd Edition

Colin Robson

# Small-Scale Evaluation

## Principles and Practice

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



First published 1999; Reprinted February, 2004.



Editor: Jai Seaman
Assistant editor: Alysha Owen
Production editor: Shikha Jain
Copyeditor: Richard Leigh
Proofreader: David Hemsley
Indexer: Author
Marketing manager: Susheel Gokarakonda
Cover design: Shaun Mercier
Typeset by C&M Digitals (P) Ltd, Chennai, India
Printed in the UK

**Library of Congress Control Number: 2017932246**

**British Library Cataloguing in Publication data**

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

ISBN 978-1-4129-6247-6
ISBN 978-1-4129-6248-3 (pbk)

# CONTENTS

# 2

# EVALUATION: THE WHAT AND THE WHY

## CHAPTER CONTENTS

# WHAT IS EVALUATION?

Dictionary definitions refer to evaluation as assessing the value (or worth, or merit) of something. The main 'somethings' focused on in this book are innovations, interventions, services or projects which involve people. For example, they might be clients of a service, or taking part in an innovation or intervention of some kind.

As explained in the previous chapter, it is commonly referred to as *program evaluation*, where 'program' is a generic term referring to any of the activities covered above. So, rather than listing 'innovations, interventions, projects, programs or services' every time, and disliking the latinate 'evaluand' for the 'something being evaluated', I'll stick with 'program' most of the time. Feel free to substitute whatever makes most sense to you in your own situation.

Consider a service which seeks to help single parents to get a job. Or a project which aims to 'calm' traffic in a village and reduce accidents. Or a program to develop healthy eating habits in young school children through the use of drama in the classroom. Or an initiative trying to cut down shoplifting from a supermarket. Or an intervention where management wants to make more efficient use of space and resources by providing bookable workstations for staff rather than dedicated offices. These situations are all candidates for being evaluated. Your own possible evaluation might be very different from any of these. Don't worry; the range of possibilities is virtually endless. Do bear your proposed evaluation in mind, and test out for yourself the relevance of the points made to it when reading through the following chapters.

Evaluation is concerned with finding something out about such well-intentioned efforts as the ones listed above. Do obstacles in the road such as humps (known somewhat whimsically in Britain as 'sleeping policemen'), chicanes and other 'calming' devices actually reduce accidents? Do children change the choices they make for school lunch? Are fewer rooms and less equipment needed when staff are deprived of their own personal space?

The answer to the first question about obstacles calls for consideration of aspects such as where one collects data about accidents. Motorists might be dissuaded from travelling through the humped area, perhaps transferring speeding traffic and resulting accidents to a neighbouring district. Also, when are the data to be collected? There may be an initial honeymoon period with fewer or slower cars and a reduction in accidents, then a return to pre-hump levels. The nature or severity of the accidents may change. With slower traffic there could still be the same number of accidents, but fewer of them would be serious in terms of injury or death. Or more cyclists might begin to use the route, and the proportion of accidents involving cyclists rise. Either way, a simple comparison of accident rates becomes somewhat misleading.

In the study of children's eating habits, assessing possible changes at lunchtime appears reasonably straightforward. Similarly, in the third example, senior management could more or less guarantee that efficiency, assessed purely in terms of the rooms and equipment

they provide, will be improved. However, complications of a different kind lurk here. Measures of numbers of rooms and amount of equipment may be the obvious way of assessing efficiency in resource terms, but it could well be that it was not appropriate or sensible to concentrate on resources when deciding whether or not this innovation was a 'good thing'. Suppose the displaced office workers are disenchanted when a different way of working is foisted upon them. Perhaps their motivation is affected and productivity falls. It may be that staff turnover increases and profits fall overall, notwithstanding the reduction in resource costs.

It appears that a little thought and consideration of what is involved will reveal a whole host of complexities when trying to work out whether or not improvement has taken place. In the traffic 'calming' example, the overall aim of reducing accidents is not being called into question. Complexities enter when deciding how to assess whether or not a reduction has taken place. With the children it might be argued that what is really of interest is what the children eat at home, which might call for a different kind of intervention where parents are also targeted. In the office restructuring example, the notion of concentrating on more efficient use of resources is itself being queried.

Such complexities make evaluation a fascinating and challenging field calling for ingenuity and persistence. And because virtually all evaluations prove to be sensitive, with the potential for upsetting and disturbing those involved, you need to add foresight, sensitivity, tact and integrity to the list.

## WHY EVALUATE?

The answers are very various. They range from the trivial and bureaucratic ('all courses must be evaluated') through more reputable concerns ('so that we can decide whether or not to introduce this throughout the county') to what many would consider the most important ('to improve the service'). There are also, unfortunately, more disreputable pretexts for evaluation ('to justify closing down this service'). This text seeks to provide help in the carrying out of evaluations for a range of reputable purposes. And to provide you with defences against being involved with the more disreputable ones.

As King and Crewe (2013) list, there has been a litany of costly national major UK policy initiatives. The post-Second World War Labour government's groundnut (peanut) planting scheme in Tanganyika, then one of Britain's African colonies, proved financially disastrous – more nuts were planted for seed than were harvested. A subsequent Conservative government wasted much more money with the Blue Streak ballistic missile system which was rejected by the military as being literally indefensible. King and Crewe bring the story up to date in a chapter on 'Blunders past and present' (pp. 25–39). Their view is that similar blunders are probably endemic in all liberal democracies. The case for thorough policy evaluation both before and at all later stages becomes compelling, given such a history.

## SMALL-SCALE EVALUATION

Few people who work today in Britain or any other developed society can avoid evaluation (and there are substantial efforts to evaluate initiatives in developing countries – in part to assess whether aid funding is money well spent). There seems to be a requirement to monitor, review or appraise virtually all aspects of the functioning of organizations in both public and private sectors. Certainly, in the fields of education, health and social services with which I have been involved, practitioners and professionals complain, often bitterly, of the stresses and strains and increased workload this involves. Many countries now have systems to evaluate the quality of university research (Hicks, 2012), although their effects are questioned (Clark & Thompson, 2016). We live in an age of accountability, of concern for value for money. Attempts are made to stem ever-increasing budget demands by increasing efficiency, and by driving down the amount of resources available to run services. In a global economy, commercial, business and industrial organizations seek to improve their competitiveness by increasing their efficiency and doing more for less. Glennerster (2012, p. R4) argues that in the UK:

> The current fiscal climate is focusing attention on the need for more efficient government. However, we have remarkably little rigorous information on which are the most cost-effective strategies for achieving common goals like delivering high quality education in deprived neighbourhoods or reducing carbon emissions.

The notion that policies and activities affecting people should be *evidence-based* is also influential in government and other circles. The increasing use of evaluation to gather evidence has been somewhat hijacked by advocates of one particular approach, the randomized controlled trial (RCT) design (see Chapter 4, p. 55). This narrowing of what counts as the best, or only admissable, kind of evidence flies in the face of how scientific understanding advances. As Rawlins (2008, p. 2159) – cited by Mullen (2016, p. 311) – points out:

> For those with lingering doubts about the nature of evidence itself I remind them that while Gregor Mendel (1822–84) developed the monogenic theory of inheritance on the basis of experimentation, Charles Darwin (1809–82) conceived the theory of evolution as a result of close observation, and Albert Einstein's (1879–1955) special theory of relativity was a mathematical description of certain aspects of the world around us. William Harvey's discovery of the circulation of the blood was based on an elegant synthesis of all three forms of evidence.

In a similar vein, Tilley (2016) makes a plea for adopting the pragmatic approach used when dealing with complex problems in engineering, a field where RCTs are virtually non-existent.

It seems highly unlikely that the future will bring about any reduction in evaluative activities, and hence there is likely to be gainful employment for those carrying out

reviews, appraisals and evaluations. While RCTs have an undoubted contribution to make, they are only one design among many. This text seeks to provide an introduction to a wide range of possibilities.

There is a positive side to evaluation. A society where there is a serious attempt to evaluate its activities and innovations, to find out if and why they 'work', should serve its citizens better. The necessary knowledge and skills to carry out worthwhile evaluations are not beyond someone prepared to work through a text of this kind.

## EVALUATION AND SOCIAL RESEARCH

A very high proportion of journal publications in the area of people-related evaluation, and of textbooks in this field, use the methods and approaches of social research. The assumption is that doing an evaluation is a type of applied social research. And that it only differs from other types of social research through its focus on assessing the worth or value of whatever is being evaluated.

The advantage of this stance is that there is a wealth of experience in carrying out social research which can therefore be called upon. Using well-tried and well-understood methods gives confidence in the trustworthiness of any findings and recommendations – and makes them more likely to be published in well-established journals.

There are disadvantages, however, not least that an extensive background and training in the use of the methods is called for. Also, that while there is a very wide variety of different methods, it precludes the use of arts-based and other non-science approaches. Insistence on the applied social science route can be particularly restrictive for those doing a small-scale evaluation. While deficiences in social science expertise in the team carrying out an evaluation can be remedied by buying in consultancy support, a lack of funding may preclude this in a small-scale evaluation.

The position taken here is that while the terms 'evaluation' and 'research' denote rather different territories, there can profitably be a considerable amount of overlap between them. A high-quality evaluation calls for a well-thought-through design and the collection, analysis and interpretation of data. Following the canons of social science research helps to ensure the trustworthiness of any findings and recommendations. This by no means restricts the practice of evaluation to those with background and training in such research, but it does mean those without can benefit from advice and support. A text such as this should help, primarily in sensitizing you to what you don't know, and hence to when you should either keep away from something or call for help.

One clear difference between research and evaluation is that the latter, if only through the derivation of the term, carries notions of assessing 'value' with it. Research, on the other hand, is traditionally seen as concerning itself with the rather different activities of description, explanation and understanding. There is now widespread recognition that

researchers themselves are not 'value-free' in their approach. However, the conventions and procedures of science provide checks and balances.

Attempting to assess the worth or value of an enterprise is a somewhat novel and disturbing task for some researchers when they are asked to make the transition to becoming evaluators. As discussed in Chapter 6, evaluation also almost always has a political dimension. Greene (1994, p. 531) emphasizes that evaluation 'is integrally intertwined with political decision-making about societal priorities, resource allocation, and power'. Relatively pure scientists, whether of the natural or social stripe, may find this aspect disturbing.

## WHAT DO THEY THINK THEY WANT?

Those involved in setting up an evaluation may be interested in achieving many different goals. 'Those involved' will vary from one evaluation to another. They are often referred to as the sponsor (or sponsors). They may have responsibilities to get new initiatives up and running – and have access to funding. Or there is a group of some kind who see the need for it. For an evaluation of an already existing program, current staff, clients and others have a legitimate interest. Table 2.1 lists several of their likely concerns, linked to some questions they might ask. The table does not provide an exhaustive list, nor is it intended to suggest that evaluations must, or should, be exclusively focused on just one of these goals. These issues are returned to in Chapter 4 which considers their implications for the design of evaluations.

In some ways the goal of 'improvement' of a program is rather different from the others. Generally, any improvement will concern changes in respect of one or more of the other goals. It could be that client needs are met more adequately, or that the outcomes are improved or efficiency increased, etc. When a program is running, it is rare to find that those involved are not interested in its improvement. Even with a well-established and well-regarded program it is difficult and probably unwise to accept it as perfect. More typically there will be substantial room for improvement, and an evaluation of a program with problems will be better received by those involved if improvement is at least one of the purposes. Indeed, there are ethical and practical problems in calling for the cooperation of program staff if you can't claim, with honesty, there is something in it for them. Potential improvement of the program is a good incentive.

It is very likely that, even when there is a clearly expressed wish to concentrate on one area, they would like to have some attention to others. For example, they may indicate the main thing is to find out whether their goals are being achieved, but say they are also interested in the extent to which the program is being delivered as originally planned.

An important dimension is whether you are dealing with an existing program or something new. If the latter, you may well find the sponsors are seeking help in deciding what this new thing might be. In other words, they are looking for some kind of assessment of the presently

unmet needs of potential clients. Perhaps they appreciate that current provision should be extended into new areas as situations, or responsibilities, or the context, change. This is not an evaluation of a program per se but of the need for a program.

With a currently running program, the main concern might still be whether needs are being met. Or it could shade into a more general concern for what is going on when the program is running. As indicated in Table 2.1, there are many possibilities. One of your initial tasks is to get out into the open not just what the sponsors think they need, but also what they will find most useful but have possibly not thought of for themselves. Sensitizing them to the wide range of possibilities can expand their horizons. Beware, though, of encouraging them to think all things are possible. Before finalizing the plan for the evaluation you will have to ensure it is feasible given the resources available. This may well call for a later sharpening of focus.

Understanding why a program works (or why it is ineffective) appears to be rarely considered a priority by sponsors, or others involved in the running and management of programs. Perhaps it smacks of the theoretical, only appropriate for pure research. However, devoting some time and effort to the 'why' question can pay dividends with eminently practical concerns such as improving the effectiveness of the program.

**Table 2.1** Likely concerns of interested parties

| To find out if client needs are met | To improve a program | To assess the outcome of a program | To find out how a program is operating |
|---|---|---|---|
| What should be the focus of a new program?<br>Are we reaching the target group? | How can we make it better (e.g. in meeting client needs; or in its effectiveness; or in its efficiency)? | Is it effective (e.g. in reaching planned goals)?<br>What happens to clients after following the program?<br>Is it worth continuing or expanding? | What actually happens during the program?<br>Is it operating as planned? |

*Note*: For 'program' read 'service' or 'innovation' or 'intervention' (or 'programme') as appropriate.

## WHAT ARE THEY GOING TO FIND CREDIBLE?

Through discussions with the sponsors and others involved in the program to be evaluated, you will be beginning to clarify what it is they want to know. It is worthwhile at this early stage also to get a feel for the kind of information they are likely to find credible. Providing such information is likely to increase the chances that the findings will be used. Some audiences value what might be termed 'hard' data: numbers and statistics. For others, the telling quotation and circumstantial detail are what really communicate, whereas tables of quantitative data merely alienate them. What you find out about this will influence both the kind of evaluation design and the way in which reports are presented.

It is possible to have more than one way of communicating the findings of the evaluation, which can be a means of dealing with different audiences.

Your own preferences and prejudices also come into this of course. However, while these, together with client preferences, will influence the general approach, the overriding influence has to be the information that is needed to get the best possible answers to the *research questions*. These are the questions that, following the discussions that you have had with sponsors, program staff and others, you need to design the evaluation to provide answers to. They are the central concerns of Chapters 4 and 5.

It is worth stressing that you have to be credible. You need to come across as knowing what you are doing. This is partly a question of perceived technical competence as an evaluator, where probably the best way of appearing competent is to be competent. Working with professionals and practitioners, it will also help if you have a similar background to them. If they know you have been, say, a registered nurse for ten years, then you have a lot in your favour with other nurses. The sensitive nature of almost all studies can mean that, particularly if there are negative findings, the messenger is the one who gets the blame. An experienced evaluator might be told 'You might know something about evaluation but obviously don't understand what it's really like to work here', while an experienced practitioner acting as an evaluator might be told 'I'm not sure that someone with a practitioner background can distance themselves sufficiently to give an objective assesment'.