Project, Programme and Policy Evaluation

Unit 1 The History, Types and Uses of Evaluation

Contents

Unit Overview

The purpose of Unit 1 is to explain what we mean by evaluation, and how and why evaluation is used in the design, delivery and assessment of public policy. Its overall aims are to introduce the notion of evaluation, and how it can be understood alongside other means of assessment. We will set public policy evaluation in an historical context, and explain the differences between projects, programmes and policies. Finally, the unit will highlight different ways in which the term 'evaluation' is used and provide some definitions to use throughout the module.

Learning outcomes

When you have completed this unit and its set readings you will be able to:

- define evaluation and place the idea alongside other means of assessing public policies and interventions
- distinguish between projects, policies and programmes and identify how evaluation of them varies
- place *ex post* evaluation in the broader context of public policy evaluation
- distinguish between different dimensions of evaluation in relation to time, scale, study design and methods.

📖 Reading for Unit 1

Colin Palfrey, Paul Thomas and Ceri Phillips (2012) Chapter 1, 'Evaluation in the political and academic Worlds' and Chapter 2 'What is meant by Evaluation?'. *Evaluation for the Real World: The Impact of Evidence in Policy Making*. Bristol UK: The Policy Press.

EVALSED (2013) *The Resource for the Evaluation of Socio-Economic Development*. Brussels, Belgium: European Commission.

1.1 Introduction

The Oxford English Dictionary defines evaluation as: '*the making of a judgement about the amount, number, or value of something*', neatly encapsulating two key elements: bringing *judgement to bear*, about something that, at least to some degree, can be *measured or valued*. You will notice that the idea is expressed as a process or action – the *making* of a judgement.

This module is about that process in the context of public policy. We will be looking at what sorts of things come to be judged, the criteria that are used to make such judgements, and their political and ideological underpinnings. The aim is to equip you to take an informed view of evaluation, to understand what it is trying to achieve, how it is approached in practice, and some of the practical issues that regularly occur.

There is a large and growing theoretical and academic literature about evaluation in its own right, and how it interacts with public policy in different ways. We will touch on this throughout the module, and your module core texts and readings contain many references you may be motivated to follow up.

The module is unashamedly practical in intent and content and the aim is to provide you with a path through what can often seem to be a complicated landscape, full of technical terms, mysterious methodologies and political dangers. It also aims to give some insights into the work of evaluators, and to encourage you to consider evaluation as a career option. As demand for the practice grows, so does the demand for evaluators, and in the career there are opportunities to study a diverse and fascinating range of public policies in action, to uncover learning that no one else has found and to influence how policies and programmes are shaped into the future.

I (your module author) have been professionally involved with evaluation for more than twenty years, and have been both the evaluator and the evaluated. I have worked on very small, local projects with limited goals and resources, all the way up to multi-billion Euro funding programmes taking place in dozens of countries. That experience has been used to inform this module, and right here at the beginning are some views I have come to which you might like to keep in mind as we progress:

- Evaluation is a highly worthwhile and rewarding enterprise. Notwithstanding the problems and compromises we will consider during the module, it is an essential part of good public policy, and the evaluator can play a key role in understanding what has happened and influencing the future for the better.
- There are rarely, if ever, clear-cut, unambiguous answers to evaluation questions. As an endeavour based on judgement, the values and beliefs of those involved play a key role in evaluation. This does not invalidate the exercise, but should to be consciously understood and acknowledged.

- Public policy evaluation is by definition a political exercise. Keep in mind where the power lies, what motivations may be in play, and who is paying for what.
- Several partial, incomplete and messy sources of information which all point to the same conclusion may well be stronger evidence than one highly specific indicator which lacks separate corroboration.
- Despite the complexity of typical evaluators' language, *ex post* evaluation in public policy is in fact conceptually rather simple. Something has been done, and the evaluation is trying to find out what happened, whether it was what was hoped for, whether it was of value, and what explains it all. Hold on to those questions irrespective of how complicated it all seems to be getting.
- The people who commission evaluations really do not want to be told in your final report that further research is needed!

Having set the scene, let us turn now to some of the history and purpose of evaluation in public policy.

1.2   Evaluation in Public Policy

Evaluation in public policy has its origins in the attempts to systematise and professionalise policy-making and programme delivery. If you have studied the CeFiMS module *Public Policy and Management: Perspectives and Issues*, you will have learned something of the history of public management and how during the 20th century there were various approaches to making it more accountable, efficient and effective, many of which depended on having a good understanding of what was happening in practice. We begin this section with two short summaries about where modern evaluation originated.

> 📖 **Readings 1.1 and 1.2**
>
> First read Chapter 1, pages 5 to 10 – the sections 'Evaluation in the political context' and 'the development of evaluation', of Palfrey *et al* (2012) *Evaluation for the Real World: The Impact of Evidence in Policy Making;* and after you have taken notes on the main points, study pages 11 to 17 of the section, 'History and purpose of evaluation', drawn from a European Union initiative.
>
> ✍ As you read this material think about what sort of evaluation exists in your own country or where you are currently working.
>
> - Does it have a long history, or is it newly arrived, perhaps brought by external donor organisations?
> - How does it fit with the different purposes that evaluation has fulfilled over its history?

Notice that both of these sources trace the evaluation of public policy to the mid-20th century in the United States, and subsequently its spread to Northern Europe including the United Kingdom, Scandinavian countries and France. In the first instance, evaluation in the public sector focused mainly

on programmes, the relatively large-scale actions organised by government and delivered by public agencies. The sources refer, for example, to education programmes, support for returning soldiers following the war, and programmes to reduce criminal behaviour.

The history of evaluation moves from these origins as an essentially managerial exercise – to help deliver the government's intentions better by looking at how effective their actions were – to a much more political and contextualised endeavour, needing to take into account multiple perspectives and interests. Governments have become more interested in evaluating the impact of their *policies*. By this we mean the overall strategic direction they are seeking to pursue, often by a number of different means, not all of which will be enacted directly by government agencies.

The evolution of evaluation also moved downwards to more localised and decentralised public *projects*. A key distinction we make throughout this module is therefore between these different levels, and the table below provides working definitions:

Table 1.1

Definition	Example
Policy	
The strategic direction decided upon by government as to the change it wishes to see and, in strategic terms, how this will be achieved. Usually the responsibility of politicians.	In the United Kingdom, policy towards unemployment and poverty currently focuses on a policy to 'Make Work Pay', in other words to ensure that people in work will always be economically better off than those unemployed and receiving state welfare support. The core idea is to make sure there are no financial incentives for citizens to remain unemployed. This is to be achieved through reforming the welfare system; incentivising employers to recruit more people; and working with unemployed people to make them more employable for the jobs available in the labour market.
Programme	
A unified group of specific actions and tools put into place to deliver the goals of a given policy. Usually the responsibility of public agencies such as ministries to organise and ensure delivery.	In the UK, as one part of the Make Work Pay policy, the Government created 'The Work Programme', giving responsibility to the national Department for Work and Pensions (DWP) for its design and delivery. This was targeted on people who had been unemployed for long periods, had failed to find work, and were continuing to claim benefits. DWP contracted on a competitive basis with private companies in different parts of the country who offered to provide training, advice and personal support to unemployed people to help them find work. The contracts stipulated that the company would be paid only once their clients had found work and remained employed for at least six months. This is only one of a number of different programmes created to deliver Make Work Pay; others focus on changing entitlement to welfare benefits and offering subsidies to employers willing to take on unemployed people
Project	
The point at which programmes are delivered in practice. Tend to be delivered at a sub-national or local	Under the Work Programme, local projects were established by the contracted organisations in the areas for which they were responsible. The projects have local offices, staffed by employment advisors, who work with local unemployed people on a one-to-one basis. Projects

Definition	Example
level, and may be done by public, private or civil society organisations, often on a sub-contracted basis.	vary in the types of support they offer, having freedom to deliver whatever they think will best get the individual into a long-term job.

It is important to understand that evaluation may be undertaken at any one of these scales, with different questions and issues likely to arise. When looking at the history of the subject you should keep in mind the evolution from managerialism up to policy justification and down to project learning.

✎ Exercise 1.1

Think about an area of public policy in your own country or one with which you are familiar, and identify one or two linked policies, programmes and projects.

- Can you trace a clear relationship between them?
- How well have they worked in their different ways to achieve desirable change?

Share your findings with fellow students online.

So far we have looked at evaluation from the point of view of individual nation states. Understanding its history in recent times however also needs to take into account the influence of the growing number of international and supra-national bodies which emerged in the second half of the 20th century. The EVALSED reading describes how the European Union has embraced evaluation as a key tool both for justifying its large-scale programmes, and for spreading good ideas across its Member States. As you study evaluation you will find that many of the ideas, norms and methods in current use have their origins in multi and supra-national bodies such as the EU, World Bank and different United Nations agencies. Autonomous national governments in wealthy countries are able, on the whole, to decide upon and then execute policies as they choose – with evaluation as part of the process of local democratic accountability and improvement of effectiveness. In the case of supra national bodies or large-scale donors providing aid or development support to poorer countries however, the dynamic can be very different. Evaluation here may well be much more to do with power relationships and the ability of donors to retain control over policies and programmes:

'an emerging challenge is the current push towards top-down processes of assessment and evaluation. In some cases these take the form of centralised results frameworks, which ignore the context of change and try to aggregate results over multiple locations in order to claim wide spread impact. For example, [the UK's Department for International Development] DfID's new business model has core results that set targets of, for example, millions of people assisted to access education, or millions of bed nets distributed to prevent malaria. While the figures provide comfort to the public that their money has been directed to helping people, the focus on simple results narrows the monitoring and evaluation agenda, pushing practitioners to assess short term tangible outcomes instead of the processes and other changes that would help assess the less tangible, but often critical, changes in the political and

> social context. It is those changes that are often required to ensure that
> the education services are directed towards poor people, and that the bed
> nets actually reach the children living in remote rural areas. Further, that
> the services will be sustained and that people have the freedom and
> ability to use the services freely. Monitoring and evaluation has to avoid
> being simplified and narrowed by what is essentially a political agenda.'
>
> <div align="right">Source: Roche et al (2012) p. 7.</div>

We have looked in some detail at the development of evaluation, and its po-
litical settings. As well as helping us understand where the idea has come
from, it is also important to consider what this means for the evaluator. On
the one hand evaluators of public policy tend to present themselves as scru-
pulously independent and design their methodologies to deliver answers
justified by the evidence. On the other, current evaluations take place in an
environment where policies are subject to intense scrutiny, with different
views as to their validity and intentions. In these circumstances the ways in
which evaluation is procured, delivered and reported may be contentious,
and the evaluator will need to be highly conscious of their place in the wider
context. It is also important to be clear about how evaluation differs from
other judgement based process, to which we turn next.

1.3 Judgement: Evaluation, Audit, Inspection, Journalism and Academic Research

We saw in the introduction that evaluation is about making judgements. In
judging public policy, evaluation is not alone: there are other processes
which also consider how public policies are performing and the effects – for
good and ill – that they have.

📖 **Reading 1.3**

Return to Palfrey *et al* (2012), and study the first sections of Chapter 2, pages 17 to 28.

✍ Note the main points raised under each section heading.

Palfrey et al (2012) Extract from Chapter 2 'What is meant by evaluation?' in Evaluation for the Real World: The Impact of Evidence in Policy Making. pp. 17–28.

In this reading you will have seen that there are different processes, profes-
sions, standards and terms used in public policy, all of which might be
considered as evaluative. Palfrey and his co-writers discuss the following:

- performance indicators
- audits
- inspections
- accreditation
- evaluation.

To these we might also add other processes which pass judgement, such as
journalism, particularly the more serious minded journalism which looks in
detail at what governments do and seeks to 'speak truth to power'; and the

academic research which takes place in many universities, independent of government, and which seeks to understand and comment upon policy and its associated programmes and practices in different ways. We can also add 'monitoring', which is often used as a joint term: 'monitoring and evaluation': M&E.

✎ Exercise 1.2

Copy the table below and then from your reading, and from your own views and experience, summarise in the table what you see as the differences between the different types of judgement you have been reading about:

Table 1.2

Evaluation
Performance indicators
Audit
Inspection
Accreditation
Journalism
Academic research
Monitoring
Any other similar terms?

Some possible answers to this question are provided, but don't check them until you have filled in the table yourself.

Share your views on the VLE. How similar or different are your views from those of your fellow students?

For the purposes of this module we will continue with the assumption that 'policy, programme and project evaluation' does exist as a distinct endeavour – if for no other reason than that policy makers and contractors of research services say they do! Our consideration of other methods of assessment shows that evaluation needs to be seen alongside other approaches, each of which can be defined separately, but that there are overlaps in relation to the methods used to collect and analyse data, the people interested in the results, and the overall intentions of understanding what has happened in practice.

1.4 Different Types of Evaluation

As mentioned in the introduction, this module is intentionally practical. Having placed evaluation in broad terms alongside other types of assessment – and acknowledged the overlaps – we are going to proceed on the basis that evaluation of public policy is a distinct endeavour, and that good quality, valuable evaluations can be designed and undertaken.

It will not surprise you to learn that there are different forms of evaluation, different terms you will encounter and different methods. This module is specifically about the kind of evaluation that looks back over public policies and programmes, to give judgements about their value, efficacy and impact, and to learn lessons – that is, '*ex post*', summative or retrospective evaluation.

There are many other terms you will encounter, including the following:

- *Ex ante* evaluation
- Interim evaluation
- Impact evaluation
- Formative evaluation
- Theory-based evaluation
- Policy evaluation
- Goals-free evaluation
- Experimental and quasi-experimental evaluation
- Economic appraisal
- Benefit–cost Analysis (or Cost–benefit analysis)
- Social cost–benefit analysis
- Social return on investment
- Project evaluation
- Qualitative evaluation
- Quantitative evaluation
- Programme evaluation.

✎ **Exercise 1.3**

Put a tick against all those terms you have encountered previously.

- What do they mean to you?
- How do they differ one from another?
- For those you have not ticked, what do you think they mean?
- Are there any other terms you've noticed that are not included here?

Tell your fellow students about them.

All this terminology is not terribly helpful for those new to evaluation, since it mixes up different fundamental elements – evaluation is being qualified by the addition of terms that relate to entirely different contexts. Annexed to the module introduction is a detailed glossary of terms to help you understand what they mean, but it is useful to have a broader framework to help in the distinction between different terms. One way to make sense of all this is to think of evaluation having four primary dimensions:

1. purpose
2. time
3. scale
4. study design.

1. The *Purpose* dimension

The most basic distinction is what the evaluation is for, its purpose. There are many fine gradations of purpose, but essentially they can be divided in two: evaluating *what* has been achieved, and evaluating *how* things have been done. You are likely to encounter the following two terms for this distinction:

Definition 1.1   Impact evaluation

Where the study is intended to show what result – usually numerically – the policy or programme either will (or more accurately may) achieve, or what is has achieved. Impact evaluation may be undertaken as part of the forecasting stage of policy design, as an *ex ante* evaluation exercise to help decide what to do; or at a mid point; or at the end.

It is intrinsically linear, based on observing or predicting what effect will result or has arisen from an identified action, once other possible causes have been taken into account.

Definition 1.2   Formative evaluation

This is where the evaluation is designed to help understand and improve on the way in which an intervention is being conducted. Rather than demonstrating impact, a formative evaluation is about how things are being done, and whether they can be improved.

Formative evaluation typically involves the evaluator in obtaining perspectives from a range of participants in a project or programme (they are rarely encountered at the policy level) on how it is performing in practice and what might be done differently. Formative evaluations typically consider internal processes and procedures, and test the extent to

which actions taken can be seen to be impacting on effectiveness and efficiency of delivery, through observation, document reviews and analysis of monitoring data.

Underlying this binary distinction are, of course, many more specific purposes: evaluation to help in decision making; to hold different parties to account; to judge between different approaches to solving a particular problem and so on. In practice, many evaluations include both purposes – they are not mutually exclusive. The key point here however is that when you encounter these terms, they are about the purpose of the exercise, rather than about other dimensions such as time, which we look at next.

2. The *Time* dimension

Policies and interventions take place over time, and they are evaluated at different points, for different reasons. A way to understand this is to use idea of the policy cycle, which follows rational steps from the identification of problems through policy formation, programme delivery and assessment of impact, and then begin the loop again with redefinition of the problem. This is illustrated in Figure 1.1, which is adapted from Hogwood and Gunn's rational policy analysis model:

Figure 1.1   Stages in the Policy Cycle
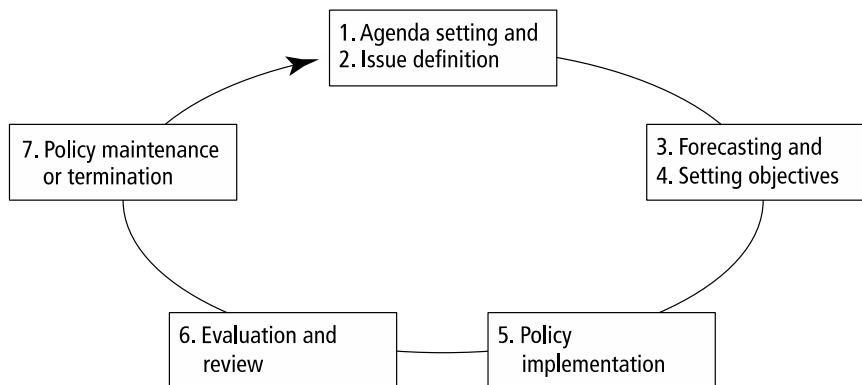
Source: adapted from Hogwood and Gunn (1984)

Evaluation crops up at different points in the cycle, and some terms are used specifically to show this, as shown in Table 1.3.

Table 1.3

Stage in the Policy Cycle	Associated time related evaluation terms
1. Agenda setting	Evaluation of all types from previous or analogous policies / programmes. Evaluation Information and conclusions are used to help understand what is already known, and identify possible courses of action
2. Issue definition	
3. Forecasting	*Ex ante* or prospective evaluation: interchangeable terms. Used to predict what might happen under different options; and to establish targets and benchmarks against which to measure progress and impact
4. Setting objectives and priorities	

5. Policy implementation	Interim evaluation: to assess progress at an interim point as to whether targets or goals are being or are likely to be met; and/or to improve performance for the remainder of the intervention
6. Evaluation and review	*Ex post*, summative, retrospective evaluation. Interchangeable terms for evaluation looking back over the policy and intervention, to assess achievement and results This is the main focus of this module
7. Policy maintenance or termination	

Source: derived from Hogwood and Gun's Rational Policy Analysis model (1984)

3. The *Scale* dimension

Scale also plays a role. Earlier in this unit we set out definitions of policy, programme and project, which move from the largest, most strategic scale at policy level down to the smallest, most local in the form of projects. You will encounter each of these from an evaluation point of view, and Table 1.4 looks further at what the distinctions are between them in evaluative terms.

Table 1.4

Policy evaluation	Tends to be used in two senses: either as an *ex ante* exercise to consider the potential merits and impacts of a given policy, or as an *ex post* review of whether a policy over time has delivered what was hoped. Whether before or after, policy evaluations tend to be particularly political in nature, and will commonly include the views of a spectrum of stakeholders, in addition to the use of statistics. We look at *ex post* policy evaluation in more depth in Unit 8.
Programme evaluation	The term programme evaluation may also apply to forward looking and *ex post*, as well as interim reviews. Programme evaluations often include large-scale statistical analysis, made possible by their application to large numbers of people or at a significant scale. Evaluators are frequently asked to evaluate programmes early on in their delivery so as to provide politicians with information (and publicity) as to the progress of their ideas, since it is through programmes that policies are enacted and therefore it is from programmes that 'good news' may emerge. The four-to-five year electoral cycles of many liberal democracies frequently lead to these evaluations being commissioned in the twelve months before an election, and for the results to be heavily used in campaigning – where they show success, or are buried if they show not much progress has been made. Programme evaluation is the main topic of Units 6 and 7.
Project evaluation	Project evaluations tend to be smaller scale and often more focused on particular issues, particularly where the projects in question are experimental – for example, to test a new approach to a problem. They are as likely to be formative (intended to improve performance) as summative (to show results), and will commonly be a requirement imposed by higher-level funders. Methodologically, project evaluations often have difficulty in providing robust statistics if they are relatively small scale, and where there are other influences, such as the macro economy, which may explain observed changes. For this reason, they are often dominated by case study and qualitative methods. Project evaluation is covered in Units 3 and 4.

4. The *Study Design* dimension

As we will consider in Unit 2, there are many different ways to design evaluation studies, with choices driven by factors such as the nature of the intervention; the purpose of the exercise; when it is taking place; the resources and data available; expediency and so on. Study design terms express the overall approach to the evaluation in question, rather than individual methodological techniques used to assemble data. Common terms you may encounter include the following.

Definition 1.3   Experimental / quasi-experimental evaluation

This is held by many evaluation theorists to be the gold standard: an approach which sets up specifically designed experiments to test a theory in practice, against a comparison or 'control' situation with the same characteristics but where the intervention is not applied. The 'quasi' sub-set refers to evaluations where a full 'case control' study is not possible, but where some direct comparison can nevertheless be made. The experimental approach is the standard in most medical research, the classic example being to test the efficacy of a new drug by giving one set of patients the drug and another a placebo, without either group knowing what they have been given, then comparing their outcomes. Outside of medicine, full or even quasi experiments are often very difficult to achieve for the evaluation of public policies. Almost by definition they will need to be set up before the intervention takes place, and report once it is complete.

Definition 1.4   Theory-based evaluation

A study design commonly used where there is little or no empirical data available and therefore no opportunity to demonstrate results on the ground. A theory-based evaluation sets out *how* the intervention was expected to cause the changes sought (the 'programme theory', to be discussed in Unit 2), and then seeks to test the validity of the underlying assumptions, rather than seeking direct evidence of impact. Theory-based evaluation may be established early on in the policy process – for example, if it is obvious that the intervention cannot be assessed by any other means – but is more often found where a policy is or has been implemented with little prior thought about how it should be evaluated. It is a valuable tool for the professional evaluator brought in to help provide answers and where it is simply too late or impractical to gather direct evidence.

Definition 1.5   Goals-free evaluation

This type of evaluation takes an open, broad view of a policy or intervention and seeks to establish what sorts of effects it has had in the round, rather than focused on a specific expectation of cause and effect. As an approach it is usually encountered either when the policy or programme had no clearly defined goals and targets; or where there appear to be significant consequences from the intervention which were not part of the original intentions but now need to be investigated. The use of this approach should not be necessary if proper thought is given to the goals of a given intervention early on, but if this has not happened it will often be a costly but necessary exercise. Where a policy has led to unexpected results, however, a goals-free evaluation is an important tool for the evaluator to use to give a full picture of what has occurred.

Definition 1.6   Participatory and neutral evaluation

An evaluation method where representatives of agencies, donors, local communities and other stakeholders work together in designing, carrying out and interpreting the

evaluation project. As method, this has the considerable merit of drawing on a diversity of perspectives, and will often yield ideas with regards to the sorts of impacts which should be investigated; ways to obtain information; possible pitfalls in carrying out the study; as well as opening up access for example to communities for data gathering. It can, however, be cumbersome, expensive and time-consuming, and carries the risk of undermining the objectivity of the evaluation exercise or embroiling it in power relationships between different parties.

Participatory evaluation can be contrasted with Neutral evaluation, which as its name suggests seeks to be entirely independent from the interests of any parties involved and take as objective and unattached a view as possible. In my view, full neutrality is impossible – there will always be some element of interest, bias, preference or agenda in any evaluation exercise. This does not invalidate it, but does need to be acknowledged in the design and presentation of results, particularly with regard to who is paying for the work. A variant on this approach is 'early neutral evaluation', where neutral third parties impart to the stakeholders their views as to the strengths and weaknesses of their respective positions relating to controversial issues. This can be helpful when the evaluation task is taking place in a contested environment.

Definition 1.7   Economic appraisal

Economic appraisal is arguably a method rather than approach, but is included here to make the point that many evaluations are fundamentally set up to understand the economics of the intervention – usually, although not always, the cost–benefit, *ie* whether the economic value of whatever impacts have been achieved offer a good 'return' for the resources spent on achieving them, or in relation to consequences of the intervention that are negative. As with most numerically driven evaluation approaches, economic appraisals need 'before and after' data – this may not necessarily have to originate before the intervention starts provided there is a clear time gap between one set of data and the next, with a change having taken place from the intervention between the two points. A key issue for economics-based evaluations is how to ascribe meaningful values to inputs or outputs that have no conventional market price – for example, the value of a forest cut down to make way for a new road.

Definition 1.8   Social return on investment (SROI) evaluation

SROI attempts to use the supposed rigour of economic evaluation approaches and apply them to social rather than financial returns, by giving a value to the social and community benefits of an intervention. This is the origin of claims such as those related to crime prevention: '*each dollar invested in Programme A is ultimately* worth *$1.80 in savings to society from averted auto thefts*' (McIntosh, 2012: p. 14.). In practice, the two approaches are not so different from each other as it may seem. Economic appraisals are increasingly obliged to price up a wider and wider range of costs and benefits (the choice of which to include and how to value them being fertile ground for differences of opinion) – in many ways the main difference is to do with what the final results are used for, and who gets to say how much non-monetised elements are worth in the calculation.

Definition 1.9   Quantitative evaluation

This is a rather ambiguous term that one could argue is about the type of data rather than study design. Quantitative essentially means pertaining to numbers – the *quantity* involved, such as the number of people who have obtained a job following participation in an unemployment support programme. When the term quantitative evaluation is used,

it typically means an assessment exercise that prioritises analysis and interpretation of statistics and numerical expression of features of the policy, programme or policy. However, this says nothing about the underlying purpose, or about the nature of study to be undertaken. If you encounter the term, it is best to understand that it simply means you will find a lot of numbers involved.

Definition 1.10   Qualitative evaluation

Qualitative evaluation is typically presented as the opposite of quantitative, meaning it is not about things that can be readily quantified, but about subjective issues, such as the attitudes of our unemployed example group towards the support they have received. While feelings do not have an intrinsic numerical expression, they can of course be captured numerically in a survey, so in reality the distinction between qualitative and quantitative is blurred. Although you will therefore encounter both terms, in my own view it is wise to be sceptical about whether any evaluation is, or indeed should be, either purely 'qualitative' or 'quantitative'. Qualitative when used in relation to research methods makes more sense – meaning, for example, data collecting through semi-structured interviewing or group discussions.

Unit 5 goes into more detail on different aspects of study design and data collection, in particular the most common approach to study design – the use of mixed methods.

We see then that all these different terms give us information about different dimensions of the evaluation. For any given evaluation assignment, a variety of terms may be relevant – few are mutually exclusive. Imagine an initiative that was intended to improve the mathematical ability of children by testing a new type of teaching in a sample of schools. The government commissions an evaluation to help it decide whether to apply the new method more widely. In this case, we are probably going to be dealing with the following:

- a purpose which means it is an *impact* evaluation (did mathematical ability improve?) and at the same time a *formative* evaluation (how did it work, so that when it is rolled out it will produce the same results elsewhere?)
- an *ex post* evaluation, since it has to look at whether what was intended did in fact happen
- a *programme* evaluation since the new approach was tested in a number of different locations
- a *quasi-experimental* evaluation, with the results achieved in the selected schools compared with others where it did not, with in all likelihood *economic appraisal* elements (how much did any improvement in performance cost?), using *quantitative* (and *qualitative*) data was there an improvement in average maths test scores? [What was the opinion of teachers about how easy the new approach was to use in practice?].

The clear implication of all this is that it is almost impossible to use just one simple term to describe what sort of evaluation is being used in a given situation. As we said back in the introduction:

Despite the complexity of typical evaluators' language, *ex post* evaluation in public policy is in fact conceptually rather simple. Something has been done, and the evaluation is trying to find out what happened, whether it was what was hoped for, whether it was of value, and what explains it all. Hold on to those questions irrespective of how complicated it all seems to be getting.

1.5 Conclusion

You have seen that evaluation began with governments looking for information to help manage their actions more effectively, and that it has grown to become both part of the political arena and a tool for democratic accountability. It has also become more pervasive, with small-scale interventions under the spotlight to harvest the lessons they offer, and for multinational donors to retain control over spending. You have also seen that evaluation is only one form of assessment of public policies and programmes, but shares some features with approaches such as audit or enquiry. Finally, evaluation itself has been looked at in its many manifestations, and some of the complexity in the terminology exposed – but also put into a framework to help to understand what it all means.

It is striking how the presentation and language of evaluation has in some ways changed and in others stayed the same over time – in 1961, over fifty years ago, the United States overseas development organisation USAID said the following about evaluation of its work:

Definition of Evaluation

The word 'evaluation' has been used in many different contexts, tending to confuse and obscure the objectives of the process. To some, the term has suggested an assessment of a program or project in terms of progress toward a predetermined goal; to others it has conveyed the impression of an inspection. A third group suggests that in addition to an assessment of progress, 'evaluation' must also include a *judgement* on the validity of the pre-determined goal. In light of the varying emphasis that has been placed on the term, valid criticisms have been made in the past that many 'evaluation' attempts have failed because clear and concise definitions of objectives were not developed prior to initiation of the process.

Within the total framework of the Agency's evaluation process, the phase emphasized in this prospectus should be directed toward these objectives:

    a)   an examination of the progress of a program or project towards pre-determined target;

    b)   an evaluation of the principal personnel involved in the implementation of the program;

    c)   an analysis of the effectiveness of a Mission' s operations;

    d)   the indoctrination of headquarters and field personnel in the new program and management concepts;

e) appropriate assistance to responsible operating officials, including the Ambassador, to insure strict adherence to the Agency's program and management concepts;

f) the development of Agency personnel, by assignment to this phase of the evaluation process, to staff in responsible positions both at headquarters and in the field.

Although any evaluation process must by its very nature contain certain negative factors, these must be minimized if the Agency's system is to prove successful.

Source: Stern (1961)

Notice the identification of ambiguity in use of the term evaluation; the primacy of progress towards pre-determined targets and of effectiveness, but also the rigidity of the institutional context and need for evaluation to contribute to indoctrination in and adherence to management concepts. This quote helps us to see that while there remains a healthy debate about what evaluation means and what it is for, we have perhaps moved on, and in the early years of the 21st century can recognise the contextual and contingent place of public policy, and hence of its evaluation.

References

EVALSED (2013) *The Resource for the Evaluation of Socio-Economic Development*. Available from: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf

Hogwood BW & LA Gunn (1984) *Policy Analysis for the Real World*. Oxford UK: Oxford University Press.

McIntosh C & J Li (2012) 'An introduction to economic analysis in crime prevention: The why, how and so what'. *Canada Research Report 2012–5*. Ottawa ON: National Crime Prevention Centre.

Palfrey C, P Thomas & C Phillips (2012) *Evaluation for the Real World: The Impact of Evidence in Policy Making*. Bristol UK: The Policy Press.

Roche C & L Kelly (2012) *The Evaluation of Politics and the Politics of Evaluation*. Birmingham UK: Birmingham University Developmental Leadership Programme Background Paper No 11. Available from: www.dlprog.org

Stern T (1961) *Prospectus for an Evaluation Process*. USAID Foreign Assistance Taskforce. Available from: http://pdf.usaid.gov/pdf_docs/PNABI188.pdf

Appendix 1.1

Ways to distinguish between different policy assessment terms:

Evaluation

Emphasises judgement, and should be based on robust, independent data sources interpreted in relation to the achievement of goals and performance.

Performance indicator

Limited and specific aspects of a policy or programme selected to demonstrate performance against predetermined expectations or targets. Usually numeric and focused on a limited set of those aspects seen as the most important.

Audit

Implies a comprehensive appraisal of key information and documentation against predetermined criteria, particularly in relation to conformity with externally established ethical or normative standards.

Inspection

Implies close scrutiny by independent parties to ensure compliance with standards, procedures or expectations.

Accreditation

Process to assess the performance of a body or individual to establish their ability to meet or deliver externally determined standards. This usually leads to the issuing of confirmation by an independent body of confirmation that a standard has been achieved.

Journalism

Independent commentary and investigation, drawing on self-selected and often partial data sources. Often framed in light of implicit or explicit ideological points of view.

Academic research

Implies independent inquiry based on establishing verifiable truths about a situation, with close attention to the robustness and the ability to replicate methods, and peer assessment of conclusions.

Monitoring

The systematic collection of data over time about specified indicators to provide the management team and key stakeholders with indications of the extent of progress and achievement of objectives and progress in the use of resources. Monitoring is not intrinsically about the making of judgements.