



## Evaluation

**Evaluation describes a set of activities that generate policy-relevant knowledge around the appropriateness, conduct and impacts of public policy interventions**

### Target and purpose

Evaluation is typically targeted at projects, programmes and policies. The term is also sometimes used to describe the assessment of individuals and organisations. Its general purpose is as follows:

- To learn about the effectiveness and efficiency of policy and programme interventions *ex post*, or to anticipate these *ex ante*, both with a view to shaping and justifying future interventions.
- To offer opportunities to a mix of stakeholders – particularly users and beneficiaries – to reflect upon the performance of those policies and programmes being evaluated and to make suggestions for improvements.
- To provide one of the bases for holding to account the actions (and inactions) of policy makers and programme managers.

### Practice

- An initial step involves defining the scope of evaluation, in terms of its purpose, its coverage and the assessment criteria to be used. This scoping is best done in consultation with a wide variety of stakeholders. In some instances, however, the scope might be mandated by law and predetermined.
- To map the ‘intervention logic’ of the programme or policy to be evaluated, a ‘logic chart’ can be created that logically links together its various goals, inputs, activities, outputs and expected impacts (outcomes). This should reflect and at the same time shape an evaluation’s coverage and assessment criteria, and should provide a hierarchy of expected impacts, ordered according to the timing of their anticipated occurrence (*e.g.* immediate, intermediate, ultimate). As ultimate impacts of policies and programmes may take several years to appear and are often difficult to measure directly, evaluation tends to focus on those immediate and intermediate impacts that can be detected more readily and somewhat earlier. These are then taken to be proxy measures of a programme’s or policy’s ultimate impacts. A logic chart is therefore a useful tool in helping to scope an evaluation. At the same time, it should not fully determine an evaluation’s scope, as there should be opportunities to capture the unintended consequences (both positive and negative) of a policy or programme that are likely to fall outside of its intervention logic.

- An important consideration in evaluation is to demonstrate a programme's or policy's 'additionality', *i.e.* to consider the extent to which desirable outcomes would have occurred without public intervention (the 'counterfactual'). There are different forms of additionality, namely:
  - i) Input additionality – the extent to which intervention supplements or substitutes for inputs provided by other means, *e.g.* the market, or by other actors, *e.g.* firms' own resources.
  - ii) Output additionality – the proportion of outputs that would not have been created without public intervention.
  - iii) Behavioural additionality – the difference in behaviour of a target population from public intervention. The concept of behavioural additionality emphasises that programmes have wider and more sustained effects than those that are most obvious to measure and that persistence of effects is of high value. Behavioural additionality concerns itself less with inputs and outputs and more with sustained changes in the behaviour of target groups, induced by contact with any stage of a programme or policy.
- A focus on additionality raises questions around the ability to accurately attribute observed outcomes to the public intervention under evaluation. Two countervailing tendencies are common here: first, the so-called 'project fallacy', whereby outcomes that are in reality cumulative and dependent upon the interaction of several factors are wholly (or mostly) attributed to the intervention under evaluation; and second, a tendency to under-estimate the effects of an intervention because of a narrow evaluation focus or because of the timing of an evaluation (where effects might not yet have occurred or have occurred so long ago that beneficiaries fail to attribute them to the public intervention). Awareness of these tendencies is important, even if the problems they create cannot be fully solved.
- Many different methods are used in evaluations, but the most popular tend to be surveys, interviews, documentary and statistical analysis, benchmarking, expert review, and case studies. All methods have their strengths and limitations, and these should be fully appreciated before proceeding with an evaluation.
- Evaluation provides one source of information among many others in shaping policy and programme management processes, and appreciating this is important in informing expectations around its usefulness. Furthermore, utilisation of evaluation results is often indirect, and some evaluation theorists and practitioners refer to the important 'enlightenment' role of evaluation, which, while difficult to account for, would seem to be extremely important. In this regard, the evaluation process, if open and participatory, can also provide useful benefits to those who participate.

## Appropriateness and feasibility

There is principled agreement that all policy interventions should be evaluated. However, debate centres around the following interrelated questions:

- When is the correct time to carry out evaluation?
- What is the appropriate level of aggregation for assessment, *e.g.* at the project, programme, policy, policy portfolio, organisational, or innovation system levels?
- What assessment criteria should be used?

These factors are often resolved in line with the knowledge needs of policy makers and/or programme managers. But *ex ante* evaluation, mid-term or continuous evaluation, and *ex post* evaluation each have important roles to play in the 'policy cycle', while different sorts of questions can be raised and different criteria used at different levels of assessment. Furthermore, in the pluralistic policy systems that characterise OECD countries, evaluation should be expected to serve the knowledge needs of a wider set of actors beyond programme managers and public policy officials. This should also have a bearing on the timing, level of aggregation and assessment criteria of any evaluation activity.

The feasibility of using evaluation depends upon the competency to carry out such assessment work. The necessary skills tend to be acquired over time so that outsourcing to specialist units or private consultants is common practice. Outsourcing also gives the appearance of providing independent assessment, though this may not necessarily be the case. A further key feasibility issue concerns the ability to utilise the process and results of evaluation in shaping future interventions. If this utilisation ability is lacking, then evaluation will be largely a waste of time and effort.

## Success factors

- Well-defined purpose, scope and assessment criteria at the outset, but with a built-in flexibility to adapt these in light of emerging evaluation findings.
- Evaluation timed to coincide with the need for information in decision making cycles while appreciating its wider enlightenment role.
- Involvement of multiple stakeholders who are addressed in evaluation's findings and recommendations.

## Risk factors

- Evaluation is carried out as ritual with little possibility or interest in learning.
- Evaluation results are used mechanistically in policy and funding decisions, with little scope for incorporating important nuances that might account for relative good / poor evaluative performance.
- Evaluation is carried out by inexperienced practitioners with the likelihood that its robustness and credibility are compromised.

## Evaluation

Periodically turning evaluation in on itself is important to assess what is gained from evaluation and how this gain could be improved. As the chief purpose of evaluation is learning, any assessment of its utility should focus upon learning processes and outcomes. Evaluations are commonly addressed to managers of programmes and policy areas, so it is natural to look there for learning and utilisation. However, other stakeholders should also be in a position to learn from and utilise evaluation. Unfortunately, efforts at examining the influence of evaluation have tended to take a rather narrow view, where typically attempts are made to trace the direct uptake of evaluation recommendations in new policy and management arrangements. This ignores the general learning and ‘enlightenment’ roles of evaluation, which are difficult to account for, yet probably constitute its main benefits.

## Further resources

European Commission (2002), *An RTD Evaluation Toolbox*,  
<ftp://ftp.jrc.es/pub/EURdoc/eur20382en.pdf>

European Commission (2006), *Smart Innovation: A Practical Guide to Evaluating Innovation Programmes*, [ftp://ftp.cordis.europa.eu/pub/innovation-policy/studies/sar1\\_smartinnovation\\_master2.pdf](ftp://ftp.cordis.europa.eu/pub/innovation-policy/studies/sar1_smartinnovation_master2.pdf)

OECD (2006), “Evaluation of Publicly Funded Research: Recent Trends and Perspectives”, a chapter in the *STI Outlook 2006*,  
[www.oecd.org/document/62/0,3343,en\\_2649\\_34273\\_37675902\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/62/0,3343,en_2649_34273_37675902_1_1_1_1,00.html)

OECD (2006), *Government R&D Funding and Company Behaviour: Measuring Behavioural Additionality*,  
[www.oecd.org/document/16/0,3343,en\\_2649\\_34273\\_36918928\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/16/0,3343,en_2649_34273_36918928_1_1_1_1,00.html)

US Department of Energy (2007), *An Overview of Evaluation Methods for R&D Programs*,  
[www1.eere.energy.gov/ba/pba/pdfs/evaluation\\_methods\\_r\\_and\\_d.pdf](http://www1.eere.energy.gov/ba/pba/pdfs/evaluation_methods_r_and_d.pdf)

US National Institute of Standards and Technology (2003), *A Toolkit for Evaluating Public R&D Investment*, [www.atp.nist.gov/eao/gcr03-857/contents.htm](http://www.atp.nist.gov/eao/gcr03-857/contents.htm)

W. K. Kellogg Foundation (2004), *A Logic Model Development Guide*,  
[www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf](http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf)