EVALSED: The resource for the evaluation of Socio-Economic Development

September 2013

Table of Contents

Welcome

Evalsed is an online resource providing guidance on the evaluation of socio-economic development. While Evalsed has a specific focus on evaluation in EU cohesion policy, it is also relevant to the evaluation of other socio-economic development tools.

Evalsed consists of two parts: THE GUIDE and a SOURCEBOOK

The GUIDE is designed primarily for decision-makers - specifically those who design and manage evaluations to enhance decision making on socio-economic development policies. It defines the role of evaluation in socio-economic development, discusses various ways to develop evaluation capacity and elaborates on evaluation approaches as well as providing guidance on how to design and implement evaluations and how to ensure their quality.

The SOURCEBOOK on evaluation methods and techniques is of particular interest to practitioners and those wishing to impart or acquire evaluation skills. It has been updated in 2013 to include new sections on theory based impact evaluation, regression analysis and beneficiary surveys. It is now organised alphabetically to facilitate online searches.

If you have any comment on any part of this Guide or the Sourcebook, or you have specific examples which you would like to have quoted, the Directorate-General for Regional Policy would welcome such contributions at regio-eval@ec.europa.eu.

History of Evalsed

Evalsed had its origins in the MEANS programme (Means for Evaluating Actions of a Structural Nature) which started in 1995 and culminated in a 6 volume survey of evaluation in 1999 (no longer in print). In 2004, MEANS was transformed and developed into a website – Evalsed (Evaluation of Socio-Economic Development). Evalsed was further developed in 2007, 2008, 2009, 2012 and, most recently, in 2013.

The 2013 review of the GUIDE takes account of the stronger results orientation agreed for Cohesion Policy in the 2014-2020 period. This was based on a reflection process between the European Commission, Member States and evaluation experts on the core concepts of outputs, results and impacts, the roles of indicators, monitoring and evaluation and the use of rigorous evaluation methods – both quantitative and qualitative. Certain changes on how the concepts and terminology is used are reflected in this version of the GUIDE.

While it would be impossible to mention everyone, the DG for Regional Policy would like to acknowledge and thank the following for their contributions over the years:

MEANS: Eric Monnier and Jacques Toulemonde, Centre for European Evaluation Expertise (C3E)

EVALSED website development (2004): Nick Bozeat of GHK, Elliot Stern, formerly of the Tavistock Institute

EVALSED developments (2009 and 2012): Alberto Martini, Director of Progetto Valutazione and Frans L Leeuw, Maastricht University, Mark Hart, Aston Business School

DG for Regional Policy: Anastassios Bougas, Andrea Mairate, Veronica Gaffey, Anna Burylo, Patty Simou, Daniel Mouqué, Kai Stryczynski, Alain Vander Velde

Introduction

Evaluating socio-economic development

The GUIDE is concerned with the evaluation of socio economic development in Europe which gives it a particular focus on European Cohesion Policy. The Structural and Cohesion Funds are organised in programmes and evaluation takes place at ex ante, interim and ex post stages. Programmes often form just one input to achieving wider policy goals and their evaluation contributes to policy evaluation. The programmes comprise many interventions and projects. Evaluation can take place at the level of the priority/intervention/project or at the level of the programme. The principles stressed in this GUIDE generally apply in socio economic programme, policy, project and thematic evaluation. Thus the GUIDE will be of use to those who have to plan, commission and manage evaluations of any socio-economic public policy, programme or intervention.

Who is this GUIDE for?

The readers of this GUIDE come from many of the different communities active in the evaluation of socio-economic development programmes. These include:

- Policy makers who have an interest in evidence to support their policies,
- Public sector managers and civil servants who may commission evaluations and would like an overview of what is available including the choices of approach and methods that they should be drawing on,
- Programme managers who will wish to incorporate evaluation results into the way they manage and plan their programmes,
- Programme partners who are increasingly involved as stakeholders in evaluations, consulted about evaluation agendas and expected to use evaluation findings,
- Evaluators, many of whom will have detailed knowledge of specific areas of evaluation but will benefit from an overview of a wider range of methods and approaches to support collaborative work with other members of an evaluation team.

Although the GUIDE itself is intended for general users and readers, rather than specialists, the sourcebook backs up the content with more detail on methods and techniques.

Why another evaluation guide?

These days we are not short of evaluation guides, textbooks and source material! As the profession and practice of evaluation has grown, a considerable library of evaluation books has been published. Whilst this literature mainly originated from North America, the expansion of evaluation in Europe - often in response to Cohesion Policy requirements - has spurred many new publications in Europe. The European Commission has published detailed Methodological Guidance[1] - on monitoring and evaluation - that is aligned with the Structural Fund Regulations. There is also a Financial Regulation which requires ex ante and ex post evaluation of all EU funded programmes. Public authorities at Member State level also publish guidance for those who evaluate national and European socio-economic development programmes and policies.

The obligations to evaluate and the guidance published by those who share responsibility for socio economic development policies change over time. Evaluation needs to be closely aligned to the circumstances in which the socio economic development is taking place and the key policy choices

[1] http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm

that need to be informed. This GUIDE is not a substitute for other sources and indeed it draws on and cross-refers where relevant to such sources. It is intended to speak to a wider audience - and to present evaluation approaches and practice in these kinds of programme and policy areas 'in the round'. Very often other sources are very specialised, addressing narrow areas of evaluation at an expert level. This GUIDE fills a gap in the market to broaden understandings of sound methods and good practice in an accessible form.

Updating MEANS

An important source of such generic guidance has been the MEANS collection - a valuable and comprehensive set of handbooks published by the European Commission in 1999. The MEANS collection became a standard text for European evaluators and enjoyed a justifiable reputation for its scope, ambition and coverage. Many aspects of that collection have stood the test of time and have been incorporated into this GUIDE. However, times have also moved on since 1999. In particular:

- There have been major changes in the world of evaluation practice, with the emergence of new evaluation tools, a heightened role for theory, new participatory and qualitative approaches (especially relevant to socio economic development) a stronger emphasis on counterfactual impact evaluation and an emphasis on the institutionalisation of evaluation.
- European policy has moved on, especially following the Lisbon Strategy and its successor, the Europe 2020 agenda.
- Enlargement of the EU in 2004 and 2007 also posed challenges for evaluation. Structural and Cohesion Funds have been introduced into public administrations with a relatively short experience of evaluation and consequently without a well developed evaluation culture.
- Experience of evaluating the Structural Funds developed over the 2000-2006 programming period and the regulatory approach for evaluation of Structural Funds in 2007-2013 introduce both a more strategic and flexible approach, requiring more emphasis on specific evaluation questions.
- The European Commission's proposals for Cohesion Policy for 2014-2020 have at their heart the focus on results. Evaluation provisions have been strengthened

Content

The Evalsed GUIDE is supported by a Sourcebook that provides more specialised and in depth material and which can be accessed and downloaded via the internet. The GUIDE is in four sections:

Evaluation and socio economic development provides an introduction to evaluation and its benefits.

Designing and implementing evaluation takes readers through practical issues in designing and implementing evaluations.

Developing evaluation capacity discusses how to develop evaluation capacity and strategies for capacity development are discussed.

Choosing methods and techniques introduces the methods and techniques of evaluation, in terms of their strengths and weaknesses - and appropriateness. Methods and techniques are discussed within a number of frameworks: different types of socio-economic programmes, different programme stages, different stages in the evaluation process and different evaluation purposes. Finally, types of data (quantitative and qualitative), indicator systems and data sources are introduced.

Each section ends with some 'golden rules' highlighting both good practice and rules of thumb that can be recommended to those who manage, commission, undertake and use evaluations. However, in general this GUIDE avoids being too prescriptive. This is partly because there is often no single right

way in evaluation and different approaches each have their strengths and weaknesses in different settings. Pragmatically also, the ideal preconditions for evaluation often do not exist - whether because of lack of data, problems of timing or availability of skills. Doing the best we can whilst still trying to improve evaluation capacity is a recurrent theme.

To support the Evalsed GUIDE a Methods and Techniques Sourcebook has also been prepared, which will be of particular interest to specialists and practitioners. A Glossary contains definitions of the terms used in the Evalsed GUIDE and Sourcebook.

Chapter 1: Evaluation and socio-economic development

The benefits of evaluation

1) Evaluations that make a difference

Investing time, money and effort in evaluation has to be justified in terms of the difference it makes to policy and programme success. Evaluation is not an end in itself. In socio-economic development the policy concern is to enhance the social and economic prospects of individuals, territories or sectors. Of course each socio-economic development intervention or policy has its own more specific rationale. Some may emphasise the development of productivity and innovativeness of small and medium sized enterprises, some the development of modern transport and environmental infrastructure, some the regeneration of inner cities, some the integration of disadvantaged groups and some the diversification of rural areas. All of these priorities and many more can be found in European Cohesion Policy programmes. However the justification for evaluation in all these cases is the same: can we apply evaluation approaches and methods in ways that will improve the quality of life, prosperity and opportunities available to citizens? To make a difference in this way requires that evaluation asks and answers questions that are useful to programme stakeholders whether they are managers, policy makers or beneficiaries.

The contribution of evaluation is potentially greatest in innovative policy areas where achieving success cannot be taken for granted and where implementation is not always straightforward. There is a need for sophisticated management and planning. When properly applied, evaluation can help make manageable some of the unavoidable uncertainties of complex situations. Socio-economic development is certainly complex and often faces many uncertainties: it is not a precise science. Choosing goals and measures, designing programmes and policies, implementing and sustaining a development dynamic, all require analysis, anticipation, establishing feedback systems and mobilising different institutions, agencies and population groups.

It is because evaluation know-how and practice has been shown to make a contribution to these processes that it has become such a key component in so many socio-economic development initiatives.

There are two important *implications* if we justify evaluation in these terms:

First, if evaluation is to be useful, usable and used, it needs to be seen as an integral part of decision making and management and indeed the entire process of democratic accountability. So a well-functioning evaluation system must be integrated into the policy cycle. This is why this GUIDE gives attention to the design of evaluation systems and the development of evaluation capacity inside public agencies and within professional and knowledge networks.

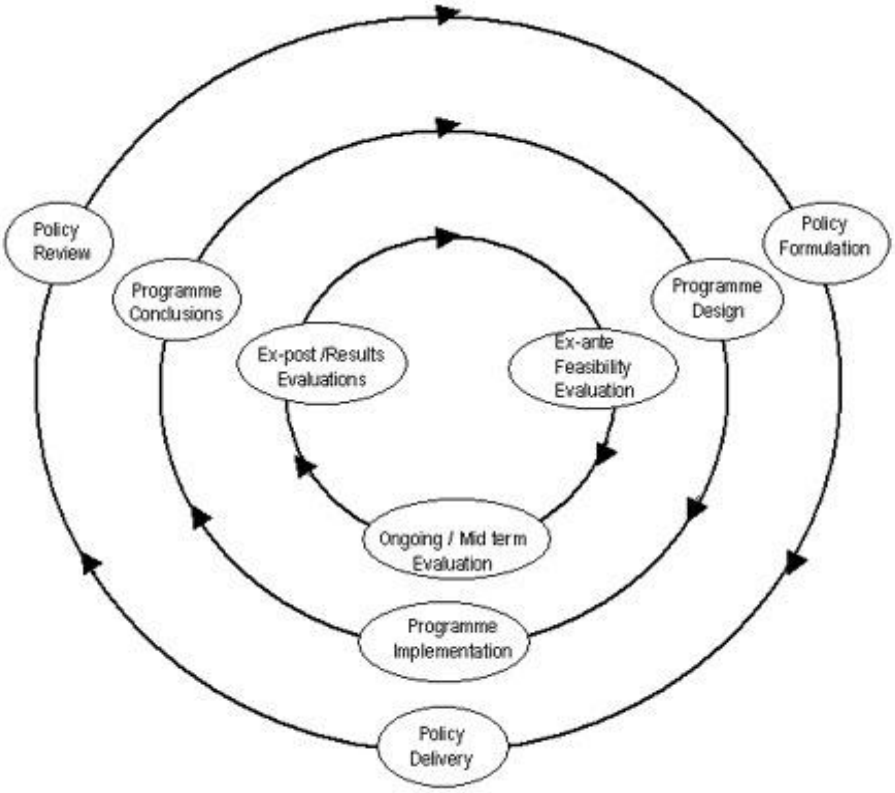
Second, evaluators and those who commission and use evaluation findings always need to balance best available methods with the demands of pragmatism. In the real world of socio-economic development we rarely have the time or resources - or the data - to implement a comprehensive state of the art evaluation. This is why this GUIDE places a strong emphasis on the kinds of strategic choices that have to be made about evaluation, for example: when are greater investments in evaluation justified? Under what circumstances are sophisticated methods needed? How can evaluation fill gaps in knowledge that in a more perfect world would have been covered before an intervention was even planned?

2) Improving policies over time

One important organising principle that runs through the GUIDE is the time-line of policy. It is common to speak of the policy cycle that begins when policy (and associated programmes) are formulated and continues through planning and resource allocation, programme design, implementation and the delivery of outputs and results. Evaluation language often follows this cycle as we can see from terms such as ex ante, interim and ex post evaluation commonly used in European Cohesion Policy. These terms are elaborated in Annex A, which also looks at the fact that programming and evaluation life cycles overlap across different programming periods..

There are three different time cycles that are important for those involved in evaluation. First, the evaluation cycle that occurs at different moments and at different stages within a second cycle, then the programme cycle which itself generates demand for these different evaluation moments. There is also a third cycle, the policy cycle which both shapes and influences programmes and inevitably also, evaluation requirements. Typically, the policy cycle is longer than the programme cycle.

Figure 1. Policy, programme and evaluation cycles



This is illustrative only, of course. There are many more stages in each cycle and they can be described in different ways. But the Figure 1 does illustrate some of the main timing problems familiar in evaluation. The inner circle moves from ex ante evaluation that documents needs to be addressed, the results to be obtained and the feasibility of planned programmes, through to interim evaluations that document progress and implementation of the different interventions and finally to ex post evaluation that focuses on results and the contribution of the programme to change. However, ex ante evaluations should feed into programme design and to policy formulation, just as interim evaluations should help shape programme implementation and deliver evidence on early effects. At the end of the evaluation cycle, ex post evaluations should contribute to policy reviews by assessing the effective use of public funds. Getting these cycles to align is desirable but does not always happen. Sometimes evaluation is of an intervention which has continued across several programming periods and the "programme" lifecycle (which delivers its funding) is not relevant. Ex-ante evaluations may be

undertaken too late to inform programme design let alone policy formulation. The results of ex-post evaluations may come in too late to inform policy reviews, but they may influence the next but one programming cycle. Changes in policy and programming can also occur when an evaluation is already underway which is not unusual in national and European programmes of socio-economic development. This can, for example, lead to changes in objectives or priorities after systems have been set up, even to the close-down of certain projects or interventions that have been the objects of evaluation. One of the advantages of involving policy makers and planners in evaluation design is to improve the alignment of these linked activities and to ensure that evidence on policy performance is built up over time and is used.

3) Designing programmes

One of the core competencies of evaluation is to gather information from different stakeholders or publics. This is especially important at the programme design stage. Ensuring the relevance of programmes to needs is essential and evaluation can contribute to this. This input on relevance is not confined to programme design at one point in time, however. In many instances of socio-economic development, programmes are kept on line by a continuous process of feedback from (potential and actual) users and from other stakeholders.

A further important role of evaluation in the design of programmes is to assist the different stakeholders to make choices on where to target their programmes. Which precise need is to be addressed? What will change if the programme is successful? Evaluation can help policy makers and planners answer these difficult questions.

4) Choosing between instruments

A well-designed evaluation system and particularly ex ante evaluation will also contribute to the selection of instruments or interventions. This may take the form of an economic appraisal that assesses the likely costs and benefits of a number of alternative instruments or perhaps large projects. It may also involve an assessment of eligibility that matches specific interventions with criteria to ensure relevance within an overall programme or regulations. An important element is an assessment of the clarity and credibility of the proposed intervention to assess the likelihood of success.

5) Improving management and delivery

A fully integrated evaluation process can make a key contribution to the way programmes are managed and delivered. By analysing monitoring data and investigating underlying causes of difficulties encountered, evaluation can provide feedback to programme management and support 'mid-course correction'. Even at an early stage there are usually early outputs, especially if there is a well-specified implementation chain and intervention logic. So evaluation of implementation brings results to the fore from the very beginning. However, many of the issues encountered at the early stages of implementation concern processes: how the parties interact, how decisions and plans are made and how new organisational arrangements and partnerships are settling down. Evaluation of such processes - even straightforward descriptions - can be helpful to all partners as well as to the main sponsoring bodies and managers.

6) Identifying outputs and results and analysing impacts

From an early stage, socio-economic development programmes need to demonstrate achievements. At the earliest stages this is likely to take the form of outputs, e.g., numbers of firms taking up a subsidy for updating their equipment or numbers of unemployed people receiving training. However policy makers are soon interested in more substantial results: firms becoming more competitive or unemployed individuals getting jobs. Only evaluation can assess impact: what is contribution of our policy to change. Evaluation should ask such questions as: has the growth of regional firms been

sustained? Or have the employment prospects of the long-term unemployed improved in sustainable ways?

For many policy makers, helping to identify, describe and quantify outputs and results is a major benefit of evaluation. To make this process really useful policy makers need to that there are clear objectives and a sensible relationship between interventions and programme goals. Programme managers, for their part, - if necessary working with evaluators - need to ensure that indicator and monitoring systems are in place and that there are clear links between the indicators chosen and the specific objectives of programmes.

7) Identifying unintended consequences and perverse effects

Even when programmes and instruments fulfil their stated objectives there will also often be unintended consequences. These can be positive or negative. For example, support for rural entrepreneurs may have spin-off benefits for urban entrepreneurs in a neighbouring city who are in the same sector or market. Sometimes such unintentional consequences can have a negative effect. For example, an instrument designed to improve the employment prospects of one group in the labour market may have negative consequences for another group. In extremis, interventions can even have a perverse effect: leading in a precisely opposite direction to that intended. For example, an intervention to promote tourism in the interest of regional development may, by misunderstanding the basis for tourist demand in a region, undermine the existing tourist trade, without creating an additional market.

To capture the results of socio-economic interventions including unintended consequences and perverse effects is essential. This is also a way in which evaluation can contribute to learning how to design programmes better and how to avoid wasteful interventions and perverse effects.

8) Levels of evaluation: policies, themes, programmes, priorities and projects

The problem of linking policies, programmes, priorities and specific interventions or projects is a perennial one in evaluation. Many good programmes do not always add up to a good policy, good programme documents do not necessarily translate into good projects and good projects do not always ensure the success of a programme. However, programme evaluation is necessarily one input into policy evaluation just as project evaluation is one input into programme evaluation.

Thematic evaluations and criteria derived from policies when applied to programme material are common ways of introducing a policy level dimension into evaluation.

There is now a tendency for evaluation to move upstream and pay increasing attention to the policy level. This reflects a willingness of policy makers to take on board evaluation results. At the same time it presents challenges for evaluators who need to view the results of their work in a wider context. Considering the policy level can also strengthen programme or priority evaluation, for example by identifying results oriented criteria for programme success.

Cohesion Policy programmes can be very large and complex covering a range of policies and areas of intervention. Good programme evaluation in such cases will depend on the availability of evaluation evidence at the level of the intervention areas, e.g., the effectiveness of aid to SMEs, the impact of improvements in a transport corridor, the effectiveness of human resource development measures. There is therefore a need for evaluation to address the policy level, while drawing a evidence from the level of the intervention area.

For the most part project evaluation, at least when the projects are of relatively small scale, is devolved to project promoters and other intermediaries. The exception to this is large-scale projects (e.g.,

infrastructure projects) that have many of the characteristics of programmes in terms of their complexity as well as size.

A common requirement for project managers and promoters is that they conduct some kind of self-evaluation. Whilst such evaluations may lack the independence considered important for external credibility they can still make an important contribution in a programme context. Requirements for self evaluation can encourage a feedback and learning culture within and amongst projects that will benefit the programme as a whole.

Those planning and undertaking evaluation need to be clear about the links between policy, programme, priority, project and thematic evaluation. The principles elaborated in this GUIDE are generally applicable to each type of evaluation.

History and purpose of evaluation

1) A short history of evaluation

Evaluation emerged as a distinct area of professional practice in the 1950s and 1960s in North America. Three strands that were most important in that early period were the evaluation of educational innovations; linking evaluation with resource allocation; and the evaluation of anti-poverty programmes. These different strands already defined some of the main evaluation traditions that continue to this day and included quantitative and experimental studies using control groups as the basis for educational testing experiments; cost benefit and economic appraisal methods; and participatory and methods involving the intended beneficiaries of programmes in the evaluation process.

Underpinning these different traditions are four main groups whose interests sometimes compete with each other in defining evaluation priorities. These include:

- policy makers, e.g., elected officials and politicians;
- professional and specialist interests, e.g., teachers in education or scientists in research;
- managers and administrators, e.g., civil servants and managers of public agencies;
- citizens and those affected by public action, e.g., the presumed beneficiaries of planned interventions.

Each of these groups makes assumptions about how evaluation can help them. For example, policy makers tend to see evaluation as a tool to ensure the accountability and justification for policy decisions; citizens are more likely to regard evaluation as a tool for democratic accountability and an opportunity to shape public interventions to their needs; managers and administrators are often concerned with the delivery of policies and programmes how well they are managed and organised; while professionals often regard evaluation as an opportunity to improve the quality of their work or even the autonomy of their own professional group.

This does not mean that evaluation - in the broadest sense the application of systematic social and economic research - was entirely absent from Europe or other parts of the world. However, it was probably strongest in Northern Europe and in those parts of Europe that had close links with the United States and Canada. From the 1970s onwards evaluation began to take root in European countries but often with distinctive traditions and emphases. In Scandinavia for example, where there is a strong commitment to democratic governance, evaluation followed in that tradition. In France evaluation mirrored the characteristics of the French state with a formal structured approach at a central government level and a more diverse and dynamic practice at regional and local levels. However, evaluation has not been static in any of these countries. For example, French evaluation

practice evolved considerably with the requirements of budgetary reform after 2000. In many countries the focus and scale of evaluative activity has reflected the changing policies of the different governments. For example, in the UK evaluation expanded considerably with the change of government in 1997.

European Structural Funds have been a major driver for spreading the practice of evaluation throughout the EU. At every stage of the programming cycle (ex-ante, interim, ex-post), there are clearly stated aims and responsibilities. It is commonly acknowledged that the introduction of evaluation into many countries in Southern Europe occurred as a result of the requirements of Structural Fund regulations. The same can be said for the Member States which joined the EU in 2004 and 2007. From modest beginnings in 1988, there is now an elaborated Structural Fund evaluation approach.

This approach includes:

- a legal obligation for programme sponsors and managers to evaluate;
- shared responsibility between different tiers of government for the overall evaluation process;
- a linked multi-stage evaluation process (ex-ante, interim, ex-post);
- the involvement of many partners in programmes and in their evaluation;
- clear links between evaluation on the one hand and programming and resource allocation on the other.

Some of the main transitions have been:

- from externally imposed evaluation obligations to internally driven demand for evaluation coming from programme managers and policy makers themselves;
- from evaluation that is bolted on to programmes at the end of a programme cycle to evaluation that is fully integrated into programmes from the beginning;
- from the expectation that evaluation results need to be disseminated largely for accountability purposes to a concern for the systematic use of evaluation throughout the implementation of a programme;
- from a view that the management of evaluation was essentially a matter of contract administration to an interest in the way evaluation can contribute to knowledge management.

Based on the experience of the 2000-2006 period, the major innovations of the 2007-2013 evaluation provisions were the introduction of the principle of proportionality and the encouragement by the Commission of an approach to ongoing evaluations based on the needs of Member States and regions. These provisions re-inforce the trend towards evaluation as a management tool to improve the design and the implementation of programmes in the overall context of accountability and a strong focus on delivering results.

For the 2014-2020 period, there has been a thorough revision of the articles concerning evaluation. There is a stronger focus on results, a much stronger emphasis on the intervention logic and concentration. Evaluation of *impact* has been introduced in the role of evaluation with requirements for managing authorities to undertake impact evaluations in line with an evaluation plan. Impact is no longer regarded as the long-term evolution of a statistic which may or may not have been influenced by policy, but the contribution of the policy to change.

The Evolution of Structural Fund Regulations:

1989-1993	1994 - 1999	2000 - 2006	2007-2013	2014-2020
EX-ANTE EVALUATION				
Ex ante assessment required to be undertaken by the Member State for each programme. Commission may undertake its own assessment.	An ex-ante evaluation must be carried out in partnership by the Commission and the Member State; it must include environmental impact.	The Member State has primary responsibility for the ex-ante evaluation. The aim of is defined; special attention must be given to the impact on the environment, the labour market and equality between the sexes.	No major change.	As for 2007-2013 with a stronger focus on the results orientation and the intervention logic.
MID-TERM EVALUATION				
No requirement	Mid term assessment required for programmes with a duration longer than 3 years, to be carried out by an independent assessor. Include are a critical analysis of monitoring data and measurement of the extent to which objectives are being achieved.	The managing authority is responsible for the mid-term evaluation in partnership with the Commission; the Commission assesses the evaluation's relevance. The evaluation is carried out by an independent evaluator by end 2003. An update of the mid-term evaluation is carried out by the end of 2005 to prepare the ground for the future. (also known as the final evaluation).	The Member State is responsible for ongoing evaluation, in consultation with the Commission. Member States are encouraged to draw up evaluation plans to guide evaluation, as the focus moves from a compliance approach to a needs-based approach. Member States should evaluate based on what they need to know and when.	An evaluation is required for each priority during the programming period to assess its contribution to its objectives. A report is required by end 2021 for each programme, summarising the findings of evaluations carried out during the programming period.

EX-POST EVALUATION				
Supposed to be carried out at national level, but not done in many cases	An ex-post evaluation must be carried out in partnership by the Commission and the Member State, to assess the impact of measures in terms of intended objectives.	The Commission has main responsibility for the ex-post evaluation in partnership with the Member State. The objective of the ex-post evaluation is defined; it is carried out by an independent evaluator within three years of the end of the programming period.	No change compared to the 2000-2006 programming period, except that it is to be completed a year earlier – 2015, the same time as spending is to finish.	No change, although the availability of the report by programme summarising evaluations will provide new material for the evaluation.
PERFORMANCE RESERVE				
None	None	By 31 March 2004, 4% of the allocation for each Member State held back at the beginning of the period is allocated to programmes whose performance the Commission, on the basis of proposals from the Member State, considers being successful. Performance is assessed based on effectiveness, management and financial implementation.	The performance reserve is optional for Member States.	Performance reserve proposed by Commission (decision to be taken by end 2013 on whether it will be applied and its scale.)

2) Different traditions and sources

Evaluation has varied roots: it is not a unified practice or derived from a single set of traditions. This is in part the result of the historical evolution, both in Europe and in North America. As already noted, it is common to highlight three important sources of evaluation thinking: the 1960s Great Society initiatives in the United States; educational innovation and curriculum innovation in schools; and budgetary control and efficiency systems. In reality these are only three sources and one could add management by objectives, participative research in community and rural development, results based management and many more.

One way of distinguishing some of these different origins is to stand back from particular examples and look at the core ideas or theories that lie behind these different evaluation traditions.

We can distinguish between four main sets of ideas:

Scientific research and methods. Many of the basic ideas and methods used in evaluation are shared with the wider research community especially in the social sciences and economics. Within the logic that combines hypotheses testing, observation, data collection and data analysis, explanations are sought for what is observed. In complex socio-economic programmes explanations are rarely straightforward. Much of the work of evaluators is an attempt to attribute observed outcomes with known inputs and vice versa.

Economic theory and public choices. Economic thinking is present within evaluation at several different levels. These include notions of efficiency and resource allocation in the face of scarcity; institutional (mainly micro-economic) incentives and behaviours; and macro-economic studies that seek to identify aggregate effects (e.g., in terms of GDP or competitiveness) of policy interventions.

Organisation and management theory. This features more prominently in evaluation as the focus has shifted increasingly to implementation and delivery of programmes and policies. This body of thinking highlights issues of organisational design, inter-organisational co-ordination (e.g., through partnerships and consortia), and issues of motivation, ownership and participation.

Political and administrative sciences. As public programmes and their managers address issues of the policy process and public sector reform they have increasingly drawn on ideas concerned with governance, accountability and citizenship. Many of the core ideas in public sector reform and the new public management such as transparency and accountability have been influenced by these perspectives.

It follows from the above that evaluators are similarly diverse. They may be economists concerned with efficiency and costs; or management consultants interested in the smooth running of the organisation; policy analysts with a commitment to public sector reform and transparency; or scientists (of various disciplines) concerned to establish truth, generate new knowledge and confirm/disconfirm hypotheses.

One of the biggest problems that those who manage or commission evaluation face is how to put together a suitable team or mix of competencies that may properly come from all these traditions.

At a systemic level (e.g., nationally or in Europe as a whole) one of the key tasks of evaluation capacity building is to build bridges between these different parts of the professional evaluation communities. Conferences, networks and professional societies that bring evaluators together are a way of increasing familiarity between those who come from different traditions as well as a way of transferring and sharing know-how, knowledge and expertise.

Despite these differences in evaluation origins and traditions it is possible to distinguish some of the main types of evaluation. These tend to cohere around two main axes. The first axis is about evaluation purposes and the second concerns evaluation methodologies.

3) The main purposes of evaluation

Evaluations always serve a broader purpose, which is to make a particular contribution to an area of public policy and its programmes. The most commonly recognised purposes of evaluation are:

- Planning/efficiency - ensuring that there is a justification for a policy/programme and that resources are efficiently deployed.
- Accountability - demonstrating how far a programme has achieved its objectives, how well it has used its resources and what has been its impact.
- Implementation - improving the performance of programmes and the effectiveness of how they are delivered and managed.
- Knowledge production – understanding what works (for whom) and why (and in what contexts.
- Institutional strengthening - improving and developing capacity among programme participants and their networks and institutions.

These various evaluation purposes are of interest to different stakeholders and also tend to be associated with different kinds of evaluation questions. For example:

- If the purpose is *planning/efficiency*, it will mainly meet the needs of planners and policy makers as well as citizens. It is these stakeholders who will be concerned with how public resources are allocated between competing purposes and deployed once they have been allocated. These stakeholders will ask questions such as: is this the best use of public money? Are there alternative uses of resources that would yield more benefit? Is there equivalence between the costs incurred and the benefits that follow?
- If the purpose of evaluation is *accountability*, it will mainly meet the needs of policy makers, programme sponsors and parliaments. It is these stakeholders that, having approved a programme or policy, want to know what has happened to the resources committed. This kind of evaluation asks questions such as: How successful has this programme been? Has it met its targets? Have monies been spent effectively and efficiently and with what impact?
- If the purpose of evaluation is *implementation*, it will mainly meet the needs of programme managers and the programme's main partners. It is these stakeholders who have an interest in improving management and delivery, which is their responsibility. This kind of evaluation asks questions such as: Are the management arrangements working efficiently? Are partners as involved as they need to be? Are programmes properly targeted in terms of eligibility? Is the time-plan being adhered to?
- If the purpose of evaluation is *knowledge production*, it will mainly meet the needs of policy makers and planners - including those who are planning new programmes. These stakeholders want to know whether the programmes assumptions are confirmed and what lessons can be learned for the future. This kind of evaluation asks questions such as: What have we now learned about what works? Are the mechanisms for intervention and change better understood? Does the logic of the programme and its assumptions need to be questioned? Is this an efficient way of achieving goals - or are there alternatives? What is the evidence on the sustainability of interventions?
- If the purpose of evaluation is *institutional strengthening*, it will mainly meet the needs of programme partners and other programme stakeholders. They will want to know how they can be more effective, how their own capacities can be increased and how beneficiaries can get the most out of what the programme promises. This kind of evaluation asks questions such as: Are beneficiaries and local communities sufficiently involved in shaping the programme and its measures? What can be done to increase participation and develop consensus? Are the programme mechanisms supportive and open to 'bottom-up' voices?

4) Learning as an overriding evaluation purpose

It is sometimes suggested that evaluation can be seen as having one overarching purpose, into which all the other purposes noted above can fit. This overarching purpose concerns learning and from this perspective evaluation has as its purpose:

> *to learn through systematic enquiry what public programmes and policies have achieved and understand how they perform in order to better design, implement and deliver future programmes and policies*.

This emphasis on learning underlines a key feature of evaluation that is consistent with the needs of socio-economic development programmes. As already observed, in these programmes knowledge is imperfect and there is a constant need to learn about different contexts and how best to combine different measures most effectively.

An emphasis on learning also highlights an important aspect of a culture of evaluation - a key element of evaluation capacity that is discussed in greater depth in Chapter 3. It is commonly agreed that for evaluation to be properly integrated into policy making there needs also to be a culture that supports learning and that is able to derive positive lessons for the future from problems or even failures as well as from success.

Methods and their roots

Chapter 4 is concerned with methods and tools (or techniques) in evaluation. Here we focus mainly on the roots or foundations that underpin these methods and tools. First five broad methodological approaches are described, then the way these connect to more general philosophical schools that are debated within most applied social and economic sciences are discussed. Many of these philosophical debates highlight the centrality of theory in evaluation which has become increasingly important in recent years. For this reason we briefly review why theory matters in the evaluation of socio-economic development and the different forms it can take. These methodological and philosophical foundations support some of the main families of methods that will be discussed in Chapter 4.

1) Methodological approaches

Looking across the different approaches to evaluation discussed above, we can distinguish five methodological approaches:

- The resource allocation approach, which is concerned with the efficient use of resources, both prospectively in terms of planning and retrospectively in terms of how resources have been used.
- The standards or target based approach, which is concerned to judge success and performance by the application of criteria.
- The explanatory approach, which is concerned to explain programme impacts and success and make causal statements about what works, when and how.
- The formative or change oriented approach, which provides positive and more complex feedback to support monitoring self correction during the life of a programme.
- The participatory/development approach, which seeks to develop networks, communities and territories through bottom-up, participatory methods.

All of these methodological emphases can be useful in evaluation: they allow us to do different things. These will be familiar to those with some experience of programme and policy evaluation. For example:

- A cost-benefit analysis that is used at the project appraisal stage would be an example of the resource allocation approach.
- An indicator study that attempts to assess whether a programme has met its objectives would be an example of the *standards or target* based approach.
- A thematic evaluation to examine the evidence across many interventions and that tries to understand what kinds of interventions to support SMEs were successful in what circumstances, would be an example of an *explanatory* approach.
- A mid-term evaluation evaluation that was intended to provide feedback so that programme managers and partners can keep on track and if necessary re-orientate their programmes, would be an example of the *formative* approach.
- A locally led evaluation intended to strengthen and build consensus among local actors, to support their agendas and increase their capacities, would be an example of a *bottom-up* or participatory approach.

One of the main characteristics of socio-economic development policies is the way they combine different interventions within a sector or territory. These interventions come from different policy areas: research and technology development; education and training; environment; infrastructure development; etc.. This is one of the distinctive challenges for evaluators of socio-economic development: how to evaluate complex sets of interlocking interventions and assess not only the contribution of each element, but the synergies between them.

Each of these policy areas brings with it its own evaluation traditions and assumptions that may be difficult to combine. There are, of course, good reasons why evaluation has taken on particular profiles in different policy areas. These differences tend to reflect specific characteristics and policy expectations in different policy areas which affect the kinds of measures and indicators that are used and the whole style of evaluation that is commonplace. For example, evaluations of science and technology interventions have tended to use bibliometric indicators as a measure of research output whilst evaluations of educational interventions depend heavily on student performance measures. In some policy areas there may be a tradition of programme managers conducting their own evaluation of the interventions that they manage. This is the case, for example, in international development where evaluation frequently consists of desk-based assessments of field projects by managers responsible for a portfolio of projects in a particular developing country or sector. In other policy areas the direct involvement of managers with a responsibility for projects being evaluated would be questioned in terms of their independence and objectivity. In the area of infrastructure, it would be usual to commission an economic appraisal of projects prior to the commitment of resources.

2) Three philosophical traditions

Three philosophical traditions underpin the broad methodological approaches to evaluation used in socio-economic development programmes.

Positivism has provided the philosophical underpinning of mainstream science from the 18th century onwards. The positivism tradition has at its heart the belief that it is possible to obtain objective knowledge through observation. Different people applying the same observation instruments should obtain the same findings. Positivist traditions aim to discover regularities and laws (as in the natural sciences). Explanations rest on the aggregation of individual elements and their behaviours and interactions. This is the basis for reductionism, whereby the whole is understood by looking at the parts, the basis for survey methods and econometric models used in evaluation. At best these methods can provide quantified evidence on the relationships between the inputs of interventions and their outcomes.

The limitations of the tradition stem from the difficulties of measuring many of the outcomes that are of interest, the complexity of interactions between the interventions and other factors and the resulting

absence of insights into what works and why. Among these limitations are: the difficulty of observing reality when what can be observed is usually incomplete and therefore needs to be interpreted by frameworks or theories; the inevitability of instrument effects, whereby what can be observed is always mediated, simplified or even distorted by the tools and techniques we use to collect data; the difficulty in most human settings to expect to find regularities and laws that do not vary across local contexts; problems of complexity where phenomena themselves change as they interact often in unpredictable ways; and the subjective and value-laden judgements of people who construct their own reality especially important in many social development settings where problems such as social exclusion are as much a matter of judgement as undisputed facts.

These limitations of positivism have led to the emergence of various post-positivist schools. The most radical, rejecting most of the assumptions of positivism, is constructivism which denies the possibility of objective knowledge. Realism, on the other hand, concentrates on understanding different contexts and the theories or frameworks that allow for explanation and interpretation.

Constructivism contends that it is only through the theorisations of the observer that the world can be understood; constructions exist but cannot necessarily be measured; facts are always theory laden; and facts and values are interdependent. In this tradition evaluators and stakeholders are at the centre of the enquiry process. The evaluator is likely to assume a responsive, interactive and orchestrating role bringing together different groups of stakeholders with divergent views for mutual exploration and to generate consensus. The evaluator plays a key role in prioritising the views expressed and negotiating between stakeholders. The stakeholder is often the most important source of data but other specific enquiries and externally generated information may be undertaken and used to inform marked differences of view.

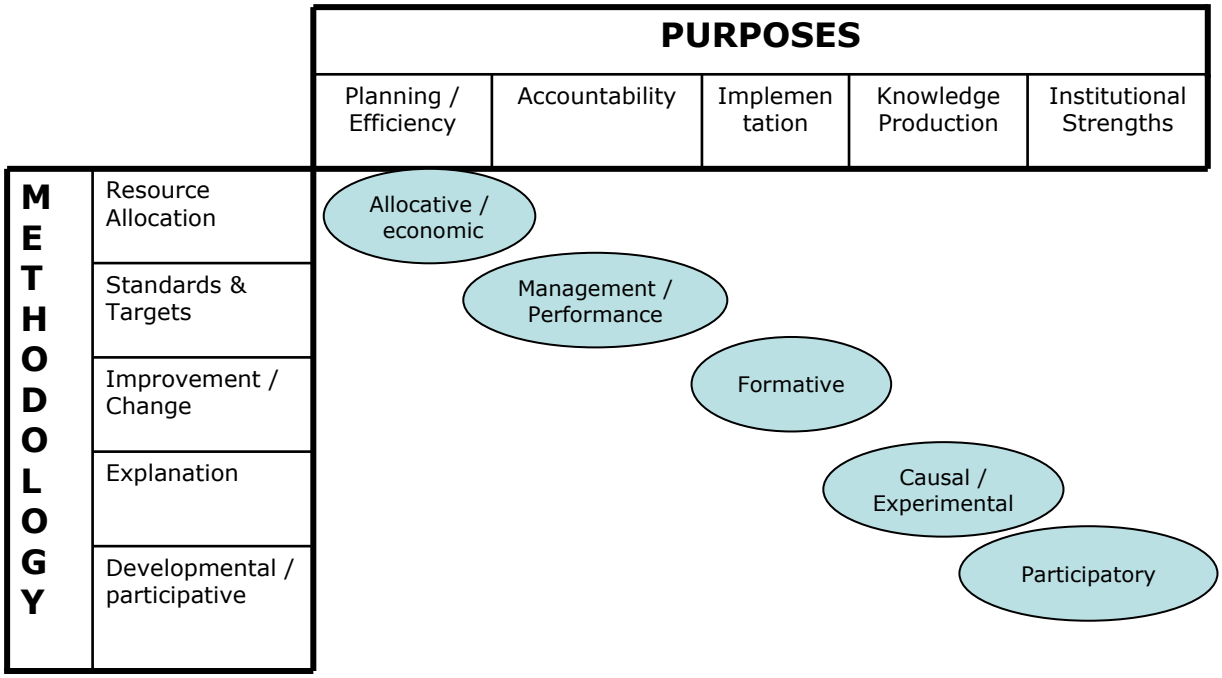
Realism seeks to open up the black box within policies and programmes to uncover the mechanisms that account for change. In doing so the tradition recognises that programmes and policies are embedded in multi layered social and organisational processes and that account should be taken of the influence of these different layers as well as different contexts. Emphasis is placed on social inquiry explaining interesting regularities in context-mechanism-outcome patterns. The systems under investigation are viewed as open. Within this tradition, the focus of evaluators is on the underlying causal mechanisms and on explaining why things work in different contexts. In depth comparative case studies are characteristic approaches of evaluation work in the realist tradition.

In practice evaluators are unlikely to see themselves as operating exclusively within any one of these philosophical traditions but will tend towards one or another depending on the circumstances of the evaluation. In general terms evaluation tools applied in the tradition of positivism will be helpful for the purposes of scrutiny. Realist approaches are likely to generate formative insights especially where the evaluation work takes place within a context where policy is being developed. Constructivist approaches can be particularly helpful in putting programmes right but are especially dependent upon the trust and chemistry between the evaluator and stakeholders.

It is important to recognise these different traditions, if only because they help explain why approaches to evaluation can be so different from each other. It is certainly useful for evaluators (and those who commission evaluations) to be explicit about their philosophical traditions and preferences.

If we combine the two axes of evaluation purpose and evaluation methodology, we can begin to identify some of the different types of evaluation that predominate in the evaluation of socio-economic development. The two axes and types of evaluation are illustrated below.

Figure 2. The two axes of evaluation, Purpose and Methodology:



At the intersection of the two axes of purpose and methodology, we have five main types of evaluation. These types cannot be rigidly located even though it helps to place them in an approximate way. For example formative evaluation can also help strengthen institutions and help management achieve targets and allocative/economic evaluations can contribute to accountability and knowledge production as well as to planning and ex-ante decisions about efficiency. However, these types do capture some of the main strands of thinking about evaluation and socio-economic development. These types of evaluation provide a useful starting point for the discussion of methods and techniques in the Choosing methods and techniques section of the GUIDE and in the Sourcebook.

One important distinction cuts across this diversity of evaluation purpose, method and type. This is the distinction between evaluations that aim to assess measure and demonstrate effectiveness and evaluations that aim to advise, develop and improve policies and programmes.

3) Three types of theory relevant to evaluation

We have already observed that theory has become increasingly important in contemporary evaluation. In part this comes from the loss of faith in pure positivism where observations were assumed to lead to knowledge and explanation independently, without interpretation. Both realists and constructivists in their different ways highlight the need for theory. But there are more practical reasons to recognise the importance of theory, following the maxim there is nothing as practical as a good theory. It is only with the help of theory that evaluation is able to analyse programme intentions and identify intervention logics; understand processes of implementation and change; or explain the partly understood effects of programmes on different policy areas. Such theories are of different types. The three most common are:

- **Programme theory**: the elaboration of Logic Models, used extensively in the context of World Bank and EU funded development programmes is one kind of simple evaluation theory that focuses on programme inputs, outputs, results and impacts. The Theory of Change is a programme theory approach concerned with opening up the black box and going beyond input output descriptions and seeking to understand the theories of actors with regard to

programme interventions and why they should work. Understanding the theories of change behind EU Cohesion Policy programmes is a priority for the design of the 2014-2020 programmes. Clarifying these theories is an important role for the ex ante evaluator.

- **Theories about the practice of evaluation**: There is a growing literature on evaluation practice, i.e., what evaluation attempts to do and what appear to be effective approaches. Such theories are the distillation of what has been learned through studies of past evaluations - for example, how to ensure that evaluations are used, how to draw conclusions from evidence, and how to put a value on a programme or intervention.
- **Theories of implementation and change**: These include understandings of: policy change; the diffusion of innovation; administrative and organisational behaviour and leadership. The theories are mainly derived from political science and organisational studies. Their application in evaluation may condition the success of programme interventions.

Evaluation to strengthen socio-economic development

1) How we evaluate depends on what we evaluate!

Socio-economic development encompasses many possible interventions including enterprise support, infrastructure, education and training, science and technology and active labour market programmes in various combinations. However, a few generalisations are possible and identifying these from the beginning prepares the foundations for thinking about and undertaking evaluations.

Most importantly if obvious: socio-economic development is about **development**. Definitions of socio-economic development are not always consistent; however, they generally encompass the following:

- Development is a discontinuous process that cannot always be predicted or controlled - it is less a matter of implementation than the initiation and management of a process of change that represents a break in what would otherwise have happened.
- A spatial dimension - all development occurs in some territory. This dimension is stronger in some programmes than others (e.g., local development programmes in urban and rural areas) but always present, especially in regional policy.
- An existing base - socio-economic development tries to build on foundations that already exist and which are seen as having further potential. This emphasises the dimension of time: development always occurs in future time, although some policy areas (such as sustainable development) may have a longer time horizon than others.
- There is a quantitative and qualitative dimension - it is about growth in numbers (of income, jobs, firms, etc.) and also the quality of work, environment, educational opportunities, etc..
- There is a policy and normative dimension - development can go in different directions and policy sets a framework of priorities and values within which choices are made.

This has consequences for evaluation. For example:

- Methods often have to be able to track change and change processes (including decisions made during programme implementation) as well as measure outcomes;
- Analysing the integration of many measures in a single territory is an important requirement;
- Ex ante assessment, pre-programming or planning evaluations need to identify resources on which development can build;
- Alongside quantitative measures, qualitative evidence of the content, standards and relevance of measures also need to be assessed;

- Policy frameworks and associated normative or value statements will define key criteria for an evaluation.

Among the most important characteristics of socio-economic development programmes, the following can be highlighted:

- They seek to address persistent and apparently intractable structural problems or fundamental needs for adaptation. So, often firms are not competitive, public administrations have limited capacities, social groups have been excluded for a long time, education and training systems are poorly linked to economic needs and the infrastructure is poor.
- They are made up of multiple interventions, intended to reinforce each other. For example, they may combine infrastructure with training, small firm development and technology transfer programmes in a single territory. Even a sectoral or thematic programme is likely to include multiple interventions or measures. This follows from an understanding that many structural problems are multi-dimensional and can only be addressed if these various dimensions simultaneously change.
- They are tailored to the needs of their settings. So business start-up programmes in the north of Germany or the south of Portugal may have similar goals but are likely to approach what they are doing in different ways reflecting the circumstances, history, resources and broader national or regional strategies in those different territories.
- They are nonetheless planned and funded within a broader national or transnational framework. Thus although tailored to particular settings socio-economic programmes are often guided by a broader concept, strategy or policy. This is so for European Cohesion Policy, shaped by a general commitment to social, economic and territorial cohesion across the EU.
- They have a strong bottom-up as well as a top down characteristic. They are designed to respond to needs and priorities of specific actors and stakeholders who may be based in a territory or sector or otherwise involved in priorities such as environmental protection or equal opportunities. These actors are regarded as partners in the socio-economic development enterprise.
- Being committed to structural and systemic change, socio-economic development programmes may not always achieve their long-term ambitions within a single programming period. Because they have long term ambitions, such policies and programmes are usually concerned with the sustainability of development outputs. It is therefore common for socio-economic development programmes to involve not only conventional outputs such as improved transport systems or new training courses. They are also likely to include associated changes in institutional arrangements and administrative capacity that will aim to ensure the sustainability of these outputs for future generations.

We can directly link these characteristics of socio-economic development with evaluation. Some of the main links are summarised below. This begins to explain why certain topics are emphasised in this GUIDE. The emphasis that is given to tracking long-term change, capacity development, the interactions and synergies between measures and to participatory methods follows directly from the characteristics of socio-economic development programmes and the assumptions that they contain.

Programme Characteristics	Assumptions that follow	Implications for evaluation
Persistent and structural development needs	Long term nature of change achieving goals will take time and require systemic change	Evaluation must capture the beginnings of long term change and put in place systems to track change over time. Evaluation should consider the wider system as well as particular outputs.
Multi-dimensional nature of programmes and interventions	Interventions are assumed to interact and reinforce each other	Evaluation must analyse the interaction between interventions. Evaluation should consider complexity and synergy
Programmes matched to settings	Programmes and interventions will differ even when goals are the same. Contexts will also differ.	Evaluation needs to consider interventions in their setting. Evaluations should assess relevance and help identify what works in different contexts. General laws will be difficult to establish.
Within a broader policy framework	Each socio-economic development programme takes forward in some way the goals of a broader framework	Evaluators can derive higher level criteria from these broader frameworks as well as from the goals of particular programmes.
Bottom-up partnerships	Bottom-up partners are always important. Sometimes they articulate needs and priorities through local/ regional knowledge and sometimes they have the dominant voice in design and implementation.	Evaluation needs to employ participatory and bottom-up approaches.
Sustainability	Programmes will include systemic change to support the sustainability of programme outputs	Evaluation should focus on those systemic changes, including capacity development - that influence sustainability

2) Theoretical underpinnings of socio-economic development

We have already noted that one area of relevant theory in the evaluation of socio-economic development programmes concerns development itself. The main sets of assumptions and theories that are used to explain and interpret the results of socio-economic programmes follow from contemporary thinking about socio-economic development.

Whereas traditionally (probably until the late 1970s), the emphasis was on managing demand through regional subsidies and other subventions (e.g., payments to the unemployed), current thinking is more directed at supply or capacity. This can take various forms, such as mobilizing underused resources, increasing the capacity and value of existing resources and transferring new resources into a region or sector. Examples of some of the most common of these assumptions and theories include:

- **Knowledge economy.** The concept of an economy characterised by the production, dissemination and application of information and know-how as products in themselves, and

the general application of information and knowledge in the production of products and services.

- **Technology transfer.** This assumes that the transfer of technologies, made possible because of the accessibility of public goods, allows less developed regions to catch up. The capacity to absorb and imitate is in the view of these theories more important than being the initial innovator.
- **Human capital.** Recognises that human resources, in particular literacy rates and education, general health and life expectancy, create conditions for increased productivity and enable people to transform their human capital into greater economic prosperity.
- **Social capital**. Again related to human well-being but on a social, rather than an individual level, through the social and institutional networks (including, for example, partnerships and associations) which support effective social action. This includes social trust, norms and networks, and political and legal systems, which support social cohesion.
- **Social exclusion.** Focuses on the disparities between individuals and communities in access and opportunities for services, jobs and infrastructure. Social exclusion impacts on balanced and sustainable economic development, development of employment and human resources, and promotion of equal opportunities. Improved economic and social cohesion is one of the EU's priority objectives and is a wide-ranging concept relating to employment policy, social protection, housing, education, health, information and communications, mobility, security and justice, leisure and culture.

One important source of ideas in structural and regional terms derives from what is often called the new economic geography. Two theories are most commonly associated with this school:

- *Local comparative advantage*. This assumes that regions have growth potential when they exploit their distinctive comparative advantage. This may take various forms: comparative advantage in trading terms (goods and services) and comparative advantage in terms of non-trading positional goods (landscape or culture, often the basis for tourism).
- *Growth poles.* That growth may require a concentration at regional level that will at first lead to an increase in disparities rather than a reduction. However, it is assumed that these disparities are subsequently eroded and growth will spread.

These different theories that underpin policy priorities have implications for evaluation. For example:

- Given the importance of additional public investment in neo-classical growth theory, whether public investment is in fact additional or whether it simply substitutes (or crowds out) investment that might have occurred otherwise becomes important to establish. As we shall see, estimating additionality, deadweight and substitution is one common element in the evaluation of socio-economic development.
- Given the importance of directing new resources to those regions and areas of investment where the potential is greatest, showing the extent to which resources are properly targeted and in fact reach their intended beneficiaries is an important activity for evaluators.

In many cases the ways in which evaluation contributes to answering big policy questions is through a qualitative understanding of what is really going on the ground. For example, many of the underpinning ideas and theories behind socio-economic development highlight the importance of technology as a resource for growth. Evaluations of a socio-economic programme or priority in this policy framework therefore need to:

- Clarify what kinds of technological investment are taking place, for instance, in skills, investment by firms in new equipment, infrastructure.
- Assess the quality of these investments in terms for example of their relevance, uptake and perceived usefulness by local and regional actors.

- Look for evidence of how efficiently, rapidly and with what degree of consensus these technological measures are being implemented.
- Identify what kinds of early direct results and consequential changes appear to be happening as a result of these inputs.
- Consider how far the innovations observed appear to be sustainable and the extent to which new capacities rather than one-off benefits are occurring.

This kind of on the ground and qualitative information can be useful for evaluation in various ways. It can:

- Help programme managers and other stakeholders better understand what is happening allowing them to improve the targeting, management and implementation of their programmes.
- Reassure policy makers that monies are being spent for the purposes for which they were made available well ahead of programme completion.
- Provide contextual information that will allow policy evaluators who are assessing the effectiveness of policy instruments more generally to interpret their findings.

The quantitative dimension may also be important for evaluation. Counterfactual impact evaluation can also contribute to answering the big policy questions for some areas of intervention (e.g., enterprise support, education and training) and where appropriate data area available. For example, the provision of support to companies to invest in new equipment could be evaluated by tracking the performance of supported companies and comparing this with the performance of an appropriately identified control group of companies not receiving support. This kind of evaluation provides more robust answers to questions about "What works?", while the on the ground and more qualitative approaches can help to answer "Why?".

3) Doing the best in an imperfect world

Both evaluators and policy makers can make over-ambitious assumptions about what evaluation can achieve in any particular programme or policy context. In an ideal world, programmes are well structured, administrative data is available, evaluations are commissioned in good time, programme promoters are clear about their objectives, the rationale for interventions has been well researched and articulated, adequate resources have been made available commensurate with the goals being pursued, the necessary skills are available to put together an evaluation team, policy assumptions remain constant throughout the life of the programme concerned and through good planning, the outputs of the evaluation will arrive in time to inform policy reviews and pre-planned reprogramming opportunities.

Unsurprisingly the real world is not like that! Policy priorities change or evolve while the programme is underway, ex-ante evaluations have not been properly conducted, programme objectives turn out to be a compromise between the conflicting priorities of different stakeholders, indicators do not capture the objectives of the programme and data are not available, and the evaluation cycle is not synchronised with the policy cycle. In these all too familiar circumstances, evaluation can still make a contribution. But this requires a twin-track approach.

First, evaluators have to be willing to produce what they can within the resources and institutional settings available, while at the same time acknowledging the limitations of what it is possible to conclude with confidence. The danger is that in response to contractual pressures, evaluators promise more than they are able to deliver or draw firmer conclusions than can be justified by the evidence available.

A second track that needs to be pursued simultaneously is to recognise the problems of resources, data, timing, programme structure, planning and the skills available at the same time as an evaluation is undertaken. On this basis, one of the most important outputs of an evaluation can be the identification of those conditions needed to improve the quality, timeliness, relevance and usability of evaluations in future. The responsibility for acting on these findings rests with those who commission evaluations and manage programmes. This is part of the task of evaluation capacity building which is discussed in Chapter 3. Of course, there will be some circumstances where the conditions are so poor that it would be unwise to conduct an evaluation. Such a conclusion might be reached on the basis of an evaluability assessment. Arguably in such circumstances the wisdom of continuing with the programme can be questioned.

Most programmes exist in less extreme circumstances, however imperfect though these circumstances might be. Nor should we underestimate the value of even small amounts of systematic information where none exists before. At the very least, the process of planning an evaluation, identifying the intervention logic, questioning the resources that are available and identifying points when evaluation outputs could inform reprogramming decisions, can help clarify thinking quite apart from the information or findings that are generated.

Golden rules

This section of the GUIDE has introduced some of the main issues for the evaluation of socio-economic development. Embedded in the various topics discussed, about the benefits of evaluation, about the nature of the evaluation task and the specific requirements of the socio-economic policy, are various hard-won good practice rules that experience has shown can help with the planning, undertaking and use of evaluation. By way of summary, these golden rules are summarised below:

1. Remember that we evaluate in order to improve programmes and policies not to undertake evaluations for their own sake. Always ask when planning an evaluation: how will the results improve the lives of citizens, the prosperity and well-being of regions and the competitiveness of economic actors. If you cannot find a positive answer to these questions, maybe you should look again at the need for an evaluation, at the very least, at the way it has been designed.
2. Aligning the time cycles of evaluations with the time cycles of programmes and policies is a worthy goal! This is the way to ensure evaluations make their maximum contribution. It is better to deliver an incomplete or imperfect evaluation on time than to achieve a 10% improvement in evaluation quality and miss the window of opportunity, when policy makers and programme managers can use evaluation results.
3. Evaluations must be integrated into programme planning and management. Programme managers need to think of evaluation as a resource: a source of feedback, a tool for improving performance, an early warning of problems (and solutions) and a way of systematizing knowledge. Evaluation is not simply an external imposition. Of course, this truism has implications for evaluators, who need to take on board the concerns of programme managers (and their partnerships) and try to take seriously their need for answers to difficult questions.
4. Evaluation is not only about looking back to rate success or failure and allocate blame. It has a contribution to make at every stage in the programme cycle. In particular, evaluation can at the earliest stage, strengthen programmes by helping to unpick intervention logics and reveal weaknesses in programme design allowing remedial action to be taken early.
5. It is no longer acceptable to gather large quantities of data in the belief that these will eventually provide answers to all evaluation questions. Data dredging is nearly always inefficient. This does not mean that data systems are not essential: they must be put in place at an early stage. However, by being clear about assumptions, by drawing on available theory and being clear about the type of evaluation that is needed, evaluations can be more focused and offer a higher yield for the resources expended.

6. Getting good work from the diverse groups which make up the contemporary evaluation professional community needs bridge and team building. Bridges need to be built at national, regional and European levels between the different traditions among evaluators, social scientists, economists, policy analysts and management consultants. So hold conferences and support professional exchange to ensure the diffusion of knowledge and know-how. This is one way of building capacity. At a micro-level, the priority is integration and the combination of different skills and competences within evaluation teams.
7. Different stakeholders, e.g., policymakers, professionals, managers and citizens, have different expectations of evaluation. If a major stakeholder interest is ignored, this is likely to weaken an evaluation, either because it will be poorly designed or because its results will lack credibility. Involving policy makers and those responsible for programmes will ensure they take results seriously and use them. Identify your stakeholders, find out what their interests are in an evaluation and involve them!
8. The policy context is an important framework within which evaluations need to be located. Of course, policy changes, or is restated in different terms and with subtly changing priorities. However, it is always necessary to keep one eye on the policy debates and decisions in order to ensure that evaluations are sensitized to policy priorities.
9. Although we have argued that all stakeholders are important (see 7 above), the emphasis on socio-economic development gives particular prominence to the intended beneficiaries of the programme interventions. Incorporating the voice of these intended beneficiaries - local communities, marginalised groups and new economic entities - in evaluations implies more than asking their opinions. It also implies incorporating their criteria and judgements into an evaluation and accepting that their experience are the justification for programme interventions. This is consistent with the logic of bottom-up, participative and decentralised approaches that are common now in socio-economic development.
10. Be pragmatic! We live in an imperfect world where resources are limited, administrators are not always efficient, co-ordination is imperfect, knowledge is patchy and data are often not available. It is nonetheless worth taking small steps, working with what is available and increasing, even marginally, the efficiency and legitimacy of public programmes. Even modest outputs can make a big difference especially when this is set within a longer-term vision to build capacity and allow for more ambitious evaluations in the future.

Chapter 2: Designing and implementing evaluation

Designing and planning your evaluation

1) Evaluation and programming

In Chapter 1 we emphasised how socio-economic development, not being a precise science, is complex and uncertain, if only because there are often several, and not necessarily mutually compatible, theories that each support different development strategies.

Planning documents are first and foremost an essential part of the cycle and a fundamental input in the policy for socio-economic development. They are, however, also policy documents, that usually have to be agreed by many actors, from different territorial levels, and with very different values, aims and priorities. It is not surprising, therefore, that they are often vague, that they try to cover every possible angle of the problem in a sometimes generic way (even if this implies spreading the available resources) thinly and that some of the objectives that they identify are mutually contradictory.

According to one vision, these uncertainties make the task of evaluating the results of socio-economic development programmes difficult, if not impossible. Without clear goals, a coherent intervention logic and theory of change, and a precise programme design - it is assumed - the identification of what to evaluate and of the criteria for evaluation becomes arbitrary and subjective. Whatever the merits of this received wisdom, there is also another way of looking at the problem. It is exactly because of the existence of multiple objectives and complex programmes that evaluation becomes essential.

Ideally, these evaluation concerns should be taken into account in the programme formulation phase, and this should help to prevent problems, such as conflicting objectives. Conceptualising the expected results in operational and therefore measurable terms (e.g., by building from an early stage a monitoring and indicator system) is a powerful means of helping decision-makers to formulate better programmes. For this reason, involving the evaluation and the evaluators as early as possible is a prerequisite for a good socio-economic development programme and for an effective monitoring and evaluation system.

Using evaluative techniques and adopting an evaluative approach from the very beginning will help stakeholders to develop a common language, while the identification of some tangible and measurable outputs and reflection on when they might contribute to results will help the general framing of the implementation process. This also allows for milestones to be established that can ensure that the process is kept on track throughout the lifetime of the programmes.

But even after the planning documents have been finalised, the contribution of evaluation to programming can be very important. At this stage evaluation can help to make sense out of a confused collection of aims and a generic shopping list of possible projects. The use of the so called logic models helps map the interests and policy priorities of the different stakeholders. If such models are developed in interaction with evaluators and policy makers, they may even lead to restructuring the programme.

It is important to note a too often forgotten truth: the value of the evaluation for policy makers lies as much in posing the right questions as in providing precise answers. (This is discussed in more detail below). Sound methodology and reliable data are very important and will yield good answers to evaluation questions. However, ultimately the added value of evaluation for decision makers consists in facing them with questions, such as: What do we want to change? Why do we believe that these policies will bring about change? Is the balance between different elements of our programme coherent? Can this level of performance be considered satisfactory?

Nonetheless, creating the conditions where evaluation questions can be answered as precisely as possible remains an important goal. In this context, the more precise the programme, the more explicit the potential trade-offs and synergies between the different goals and objectives, the more stringent the programme logic, the more reliable the programme theory (i.e., the causal links between the projected actions and the expected results), the clearer the indicators system, the easier and more precise will be the evaluation. Such an ideal situation will make the judgements about the success or failure of the programme sound, reliable and useful. It will increase accountability to the governance system and develop a better understanding of the ways in which the general goal of sustainable socio-economic development can be attained.

From the perspective of a prospective evaluator there is benefit where:

- The objectives are as explicit as possible and the concepts referred to are defined and commonly understood.
- Objectives are linked to interventions (or groups of interventions) and measurable outcomes.
- If interventions have multiple objectives, some explicit weight is attached to each objective.
- The objectives incorporate targets the achievement of which can be measured. The targets should have explicit timescales and should have been the result of negotiation between policy makers and those responsible for implementation.
- There are incentives that unambiguously encourage the achievement of objectives.

These conditions help the evaluation process to define criteria against which to judge whether an intervention might, is en route to contribute or has contributed to a successful outcome. A further condition which could support rigorous impact evaluation is reflection at the programme design stage on how the evaluation can be done. What method should be used – theory based or counterfactual or both? What data will be available? Can the counterfactual can be estimated in a quantified or qualitative way, i.e., how can we evaluate what would have happened in the absence of the intervention.

In practice these conditions have been uncommon to date and as is elaborated below, the evaluation process itself must often contribute to the defining of objectives and criteria that reflect the ambitions that were set for the interventions. However, reflection on these questions should start from the beginning of a programme in order to build costs effective monitoring and evaluation systems.

2) Planning evaluation

This section more precisely discusses various issues arising when planning evaluation:

- Scoping and defining the object of evaluation;
- Identifying and involving stakeholders;
- Analysis of the intervention logic and the theory of change and policy objectives, which underlie the interventions.

Defining the object of evaluation

The decision to evaluate is an opportunity to define limits in terms of institutional, temporal, sectoral and geographical dimensions. This is what is known as the scope of the evaluation or the "evaluation object". Defining the scope of an evaluation amounts to asking the question: What is going to be evaluated and when?

Evaluation scope can be specified in at least four respects:

- institutional (European, national or local level);

- temporal (time-period under consideration);
- sectoral (social, industrial, environmental, rural, etc.); and
- geographical (which part of the European territory, which region, town, nature reserve, etc.).

A programme is notionally delimited by finance, by the territory concerned and by the programming period. It is, however, useful to first consider:

- Is the intention to limit evaluation to the funding of the programme or to include other national, regional or local funding that is, to a greater or lesser degree, directly related to the programme?
- Is the intention to limit the evaluation to interventions in the eligible area or to extend observations to certain neighbouring areas that encounter similar development problems?
- Is the intention to limit the evaluation to funding allocated within the programming cycle under consideration or to a certain extent to include funding of preceding cycles?
- Is the intention to limit the evaluation to a part of the programme, thus allowing more in-depth analysis?

It is normally helpful to adopt a relatively strict definition of the scope of the evaluation. Experience has shown that during the evaluation process stakeholders may wish to examine almost everything. In order to reach conclusions, the evaluation should be confined to an examination of the programme or a part of the programme and its most essential interdependencies with other public policies and interventions.

This risk of the scope widening is particularly great for ex ante evaluations. These can turn into exercises in forecasting or speculation that are far from the object of the evaluation. In ex ante evaluation it is best to limit the scope of the evaluation strictly to the programme proposals.

Commissioners of evaluation are often unwilling to restrict the scope of the evaluation questions they expect to cover. One contribution that evaluators can make is to identify those questions most central to programme success based on programme theories that identify intervention logics and implementation chains. Sometimes the best way to prioritise evaluation questions and focus the evaluation is to discuss practical constraints like time and resources.

For an evaluation to be useful, the decisions likely to be taken and which can be informed by the evaluation, must be stated as precisely as possible. Often commissioners, not wanting to influence the evaluation team too much, are reluctant to express in advance the changes they think should be made or their doubts concerning the effectiveness of a particular action. The intention is commendable: reveal nothing in advance to see whether the evaluation team reaches the same conclusions! Experience shows, however, that evaluation has little chance of documenting intended decisions if these are not known in advance by those who are going to collect and analyse data in the field. Socio-economic reality is highly complex and the evaluation team is confronted with a large number of observations and possibilities for making improvements. Verifying hypotheses which are in the event of little interest to officials, managers or other stakeholders is not realistic.

Identifying and involving stakeholders

As we have already seen socio-economic development includes several different types of projects, programmes and policies - this implies the number of actors or interested parties is often quite large. Evaluation experience suggests that this is far from being an obstacle to a good evaluation. On the contrary it offers opportunities that should be exploited in order to pose the most appropriate questions and give the most useful answers.

Other factors which have reinforced the trend towards involvement of large and diverse groups of institutions and actors include the emergence of multi-level governance and application of subsidiarity, and the recognition of the role played by social capital in socio-economic development.

The emergence of local and territorial development, where different policy sectors and sources of financing are integrated in an attempt to enhance the socio-economic development of an area, makes the identification of stakeholders and their involvement in the programme formulation process (the bottom up approach to planning) an essential step of the whole exercise for certain evaluations.

Even in simpler programmes and projects there are always a number of actors whose interests are affected, positively or negatively, by the planned or implemented activity.

In all cases therefore, identifying the potentially affected actors (in ex ante evaluations) those actually affected (in interim or ex post exercises), and somehow involving them in the evaluation process is paramount to take into consideration points of view, indirect effects or unintended consequences that can be very significant for describing the effects, understanding the causality chains and judging the results. A further concern is to involve those not affected by the intervention although eligible for support.

An additional rationale for identification and involvement of stakeholders is that evaluators and programme managers have an interest in ensuring that there is ownership of evaluation findings. Only in this way is it likely that those involved will take evaluations seriously and act on recommendations or define their own action priorities on the basis of findings.

The first question that must be asked, after the scope of the evaluation has been defined, is therefore quite straightforward: Who are the individuals, the groups or the organisations who have an interest in the intervention to be evaluated and can be interested in the process or in the results of the evaluation itself? This phase in evaluation parlance is called the identification of the stakeholders.

The second question that should be asked is: what are the relevant inputs to the design, management or content of the evaluative exercise that these stakeholders can make? The involvement of the stakeholders can take place at very different levels:

- At a minimum the evaluators should make sure that stakeholders provide evidence (data, information, judgements, etc.).
- Stakeholders can be involved in steering the study, including defining priorities, evaluation questions and associated criteria, although here care must be taken that certain stakeholders do not influence too much the conclusions of the evaluation.

In practice the involvement of stakeholders in most evaluations falls somewhat in the middle. If the participation of stakeholders in the Steering Committee is restricted to the official institutional and social partners, some way to provide feedback to other actors that are able to provide information and judgements is widely practiced through the dissemination of reports, ad hoc meetings and similar instruments.

Programme theories and logic models

The stakeholder consultation phase also provides an opportunity to reconstruct the logic of the programme prior to its launch. As we have seen in Chapter 1, there are different and often competing theories underpinning interventions.

Ideally, every programme or policy would state clearly the set of assumptions on the basis of which the intended outcomes - in our case some aspect or aspects of socio-economic development - can be influenced through the resources allocated and the interventions funded. These assumptions would be consistent with each other and would be supported by evidence. To-date, this is rarely the case in practice, especially in the complex world of socio-economic development.

A further step in evaluation planning, therefore, is to reconstruct the theory underpinning the object of evaluation. This is to assess the ability of the programme to contribute to its intended outcomes. A clear identification of the reasons why it should is an important precondition to posing the right evaluation questions.

This emphasises how programming and evaluation are interrelated. Programme managers and planners need to be aware that there are tools available that can help reconstruct the chain that links objectives of the programme, the interventions funded, the outputs achieved and finally the contribution of the interventions to these results. Chapter 4 discusses various tools and techniques (logic models, log frames, programme theory, theory of change) that can assist in the reconstruction of programme intervention logics and implementation chains.

In conjunction with stakeholder consultation and analysis, the application of these methods can help to pinpoint the possible critical aspects of the programme implementation and therefore to focus the evaluation appropriately.

3) Defining evaluation questions and criteria

Defining evaluation questions

Defining evaluation questions is an essential part of the start-up of any evaluation. Evaluation questions can be at different levels. They can be:

- *Descriptive* questions intended to observe, describe and measure changes (what happened?)
- *Causal* questions which strive to understand and assess relations of cause and effect (how and to what extent is that which occurred attributable to the intervention?)
- *Normative* questions which apply evaluation criteria (are the results satisfactory in relation to targets?)
- *Predictive* questions, which attempt to anticipate what will happen as a result of planned interventions (will the interventions to counter unemployment in this territory create negative effects for the environment or existing employers?)
- *Critical* questions, which are intended to support change often from value-committed stance (how can equal opportunity policies be better accepted by SMEs? or what are the effective strategies to reduce social exclusion?)

Ideally, evaluation questions should have the following qualities:

- The question must correspond to a real need for information, understanding or identification of new solution. If a question is only of interest in terms of new knowledge, without an immediate input into decision-making or public debate, it is more a matter of scientific research and should not be included in an evaluation.
- The question concerns a result or a need. That is to say, it concerns, at least partly, elements outside the programme, notably its beneficiaries or its economic and social environment. If a question concerns only the internal management of resources and outputs, it can probably be treated more efficiently in the course of monitoring or audit.
- The question concerns only one judgement criterion. This quality of an evaluation question may sometimes be difficult to achieve, but experience has shown that it is a key factor in the usefulness of the evaluation. Without judgement criteria clearly stated from the outset, the evaluation rarely provides conclusions.

Finally it is noteworthy that not all questions that evaluation commissioners and programme managers ask are suitable to be evaluation questions. Some are too complex, long term and require data that is not available. Other questions do not even require evaluation but can be addressed

through existing monitoring systems, consulting managers or referring to audit or other control systems.

Evaluation criteria

Evaluation questions that include judgement criteria fall primarily into one of the following four categories:

- Those related to relevance;
- Those related to effectiveness;
- Those related to efficiency; and
- Those related to utility and sustainability

Figure 3: Intervention Logic

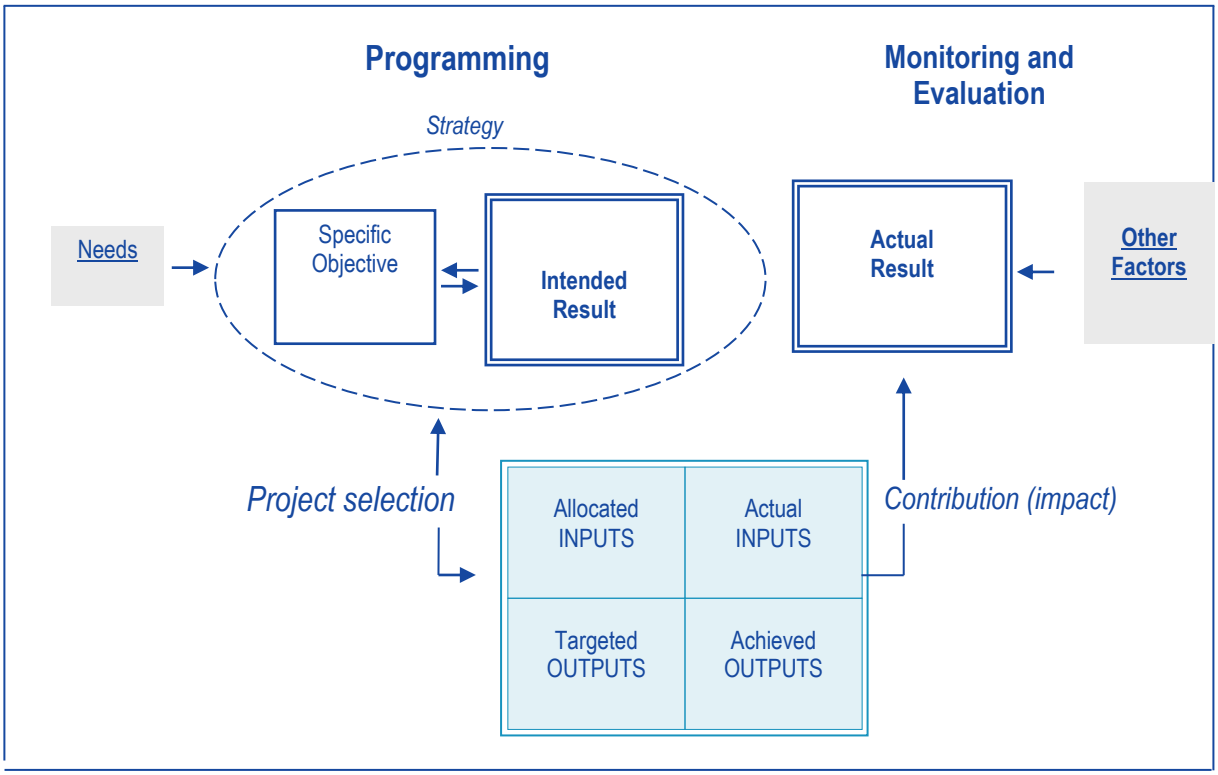


Figure 3 shows the intervention logic at programming stage and then what is monitored and evaluated. The evaluation questions and their judgement criteria relate various elements of this logic to each other. Relevance, in the context of an evaluation, refers to the appropriateness of the explicit objectives of the programme in relation to the socio-economic problems it aims to address, the needs. In ex ante evaluation, questions of relevance are the most important because the focus is on choosing the best strategy or improving its quality. In interim evaluation, the aim can be to check whether the socio-economic context has evolved as expected – what has the actual result been and how have other factors contributed? - and whether this evolution calls into question a particular objective.

The term effectiveness concerns whether the objectives formulated in the programme are being achieved – what are the actual results? What have been the successes and difficulties, and how appropriate have the solutions chosen been and what is the influence of external factors that come from outside the programme. We clarify here that impact is the change which can be credibly attributed to an intervention. In the past, there was some confusion in the evaluation debate is that

"impact" was also used (and still is in some cases) to mean longer term effects, including those which are indirect and unintended.

The term efficiency is assessed by comparing the actual outputs and the inputs – the resources mobilised. (The terms economy and cost minimisation are sometimes used in much the same way as efficiency).

The basic questions of interim evaluations and, more particularly, of ex-post evaluations, concern the effectiveness and efficiency of the interventions implemented.

The criterion of *utility* judges the results obtained in relation to broader societal and economic needs. Utility is a very particular evaluation criterion insofar as it makes no reference to the official objectives of the programme. It may be judicious to formulate a question of utility when programme objectives are badly defined or when there are many unexpected impacts. This criterion must nevertheless be used with caution to ensure that the evaluation team's selection of important needs or issues is not too subjective. One way of safeguarding against this risk is to involve other stakeholders, and in particular, intended beneficiaries in the selection of utility questions.

The term *sustainability* refers to the extent to which the results of the intervention are durable. Often evaluations consider the sustainability of institutional changes as well as socio economic impacts. (The criterion of sustainability is also linked to the concept of sustainable development which can itself be regarded as one definition of *utility*, particularly if sustainable development is defined as concerning the maintenance of human, productive, natural and social capitals rather than just the maintenance of the environment for future generations).

Typical evaluation questions relating to the main criteria are given below.

Evaluation questions related to the main evaluation criteria

- *Relevance:* To what extent are the programme objectives justified in relation to needs?
- *Effectiveness:* To what extent have the objectives been achieved? Have the interventions and instruments used produced the expected effects? Could more effects be obtained by using different instruments?
- *Efficiency:* Have the planned outputs been achieved at the lowest cost?
- *Utility:* Are the expected or unexpected effects satisfactory from the point of view of direct or indirect beneficiaries?
- *Sustainability:* Are the results including institutional changes durable over time? Will they continue if there is no more public funding?

These criteria are not exclusive and not all questions will be answered in all evaluations. It is important to be selective so as not to overburden the evaluation. Other criteria such as equity, coherence, synergy, reproducibility are also often used in evaluation. Even though programmes have their own logics and objectives, they are embedded in policies that define a broader purpose. They may ultimately be seen as contributing to social inclusion or greater economic competitiveness even though their immediate goal is vocational training or new business start-up.

Evaluators must also take seriously effects that were not part of programme design. Among the results of a programme that go beyond formal goals that evaluators should consider are:

- The experiences and priorities of intended beneficiaries who have their own criteria for effectiveness that may not accord with those of programme designers and policy planners;
- Perverse effects that are not simply unintended but may actually run counter to programme intentions reducing opportunities rather than increasing them, exporting jobs rather than creating them; and

- Results suggested by other research and evaluation, possibly drawing on theories of socio-economic development or comparative experience in other countries.

This then is an argument for evaluation not to be exclusively goal oriented but sometimes to stand aside from the logic of programmes and adopt an independent and even critical stance. This is not, however, a justification for ignoring programme goals, rather an argument to go further in order to understand the dynamics of the programme.

One set of concepts that are commonly applied in evaluations derives from economic theory and includes:

- Additionality, was the intervention additional to what would otherwise have taken place?
- Deadweight, did the intervention generate outputs and results that would in any case have occurred?
- Displacement, did the intervention cause reductions in socio-economic development elsewhere?

Evaluation can best contribute to answering questions about deadweight and displacement when the scale of an intervention or programme is large in relation to other possible explanations of outcomes. This may not be the case in smaller interventions.

Fundamentally, however, the evaluator seeks to understand the difference the programme or intervention makes and behind all evaluations is the counterfactual question: what would have happened in the absence of the programme or intervention? Of course, this question can never be answered with certainty because we cannot construct a parallel reality. There are various methods to estimate the counterfactual, using quantitative or qualitative methods which are discussed in detail in the section of the Sourcebook on impact evaluation.

Evaluability of evaluation questions

Once the evaluative questions have been identified, their evaluability has to be considered. A prior assessment has to be made of whether the evaluative questions are likely to be answerable, given available data. Will the evaluation team, with the available time and resources and using appropriate evaluation tools, be able to provide credible answers to the questions asked?

For each evaluative question one needs to check, even very briefly:

- whether the concepts are stable,
- whether explanatory hypotheses can be formulated,
- whether available data can be used to answer the question, without any further investigation,
- whether access to the field will pose major problems.

A number of factors can make a question difficult to answer: if the programme is very new, if it has not yet produced significant results or if there is no available data or the data is inappropriate. These reasons may lead to the decision not to undertake the evaluation, to postpone it, or to ask more realistic questions.

Important considerations at the evaluability stage are the probabilities that evaluation results will be obtained and used. Questions that are relevant therefore include:

- Will the conclusions be used? By whom? For what purpose? When?
- Is it politically appropriate to perform such an evaluation at this particular time or in this particular context? Is there a conflictual situation that could compromise the success of the exercise?
- Has a recent study already answered most of the questions?

4) Choosing evaluation methods, responsibilities and allocating resources

Choosing methods

Evaluation questions can be answered in a variety of ways. The choice of the method is therefore critical in order to get the most from the evaluation resources available. This is normally an operational choice that can be finalised only when the field of analysis has been reconstructed and there is enough information about the availability of data. However during the planning phase, some choices should be made, influenced by:

- the reliability of the programme theory;
- the level of consensus between the stakeholders;
- the type of programme to be evaluated;
- the point in the programme cycle at which the evaluation takes place;
- the theme or sector of intervention of the programme.

Chapter 4 provides further information and guidance on the choice of methods. The Sourcebook elaborates different methods and techniques, while the Glossary provides definition of tools in less common usage.

The role of Guidance

The fact that Cohesion Policy evaluations are sometimes compulsory and that both the European Commission and National authorities issue guidelines about when and how to perform the evaluative exercises is a mixed blessing. On the one hand it can routinise the decision to evaluate. Evaluation can become an obligation to humour the above-mentioned institutions, with no clear added value for programme managers. On the other hand it can provide a much needed guidance both to the planning authorities and to the evaluation teams about the expected behaviours and results. Certainly the presumption that evaluations should be undertaken and the availability of guidance on their scope has been an important stimulus for the development of evaluation capacity as discussed further in Chapter 3.

For the 2007-2013 programming period, the obligations for evaluation during the programming period were reduced, in an attempt to move from a compliance-based approach to one based on needs. The Commission's guidance advises the establishment of a multi-annual evaluation plan to guide the evaluation process. This approach requires that programme authorities reflect more deeply than perhaps they have done in the past on what they want to evaluate and when. This guide should provide some additional support in this regard.

In 2014-2020, the proposed approach to evaluation is strengthened but the needs based approach to evaluation plans continues with a requirement that each priority should be evaluated at least once during the programming period to assess of objectives are being met. New guidance has been developed by the European Commission.

Key Resource decisions

Evaluations can also be seen as an attempt to second-guess programme managers' choices. More often than not they are under the impression that they already know most of the things that the evaluators will tell them.

This is why it is important to involve the political authority, or at least the top management together with the most important external partners of the programme, in the planning of the evaluation. This does not mean involving them in the more technical decisions but making sure that they have the possibility to influence the following four fundamental questions:

- The reasons for the evaluation?
- Who is in charge of the overall exercise?
- How much to spend for the study?
- Who will perform the work?

Reasons for the evaluation

This is the most fundamental question. As we have seen, there are different possible general purposes of an evaluation, there are different specific aims and there are different possible evaluation questions. Making sure that the choice reflected in the terms of reference is shared by top decision-makers and by the most relevant partners of the programme lends credibility to the whole exercise.

Who is in charge?

This involves:

- the members of the Steering Committee;
- those who write the terms of reference; and
- those who act as a liaison between the administration and the evaluation team.

Those in charge must be senior enough to have direct access to the policy makers in order to share with them the knowledge that the study will produce. They must also be conversant with the theoretical and methodological problems of evaluative research. This is essential in order to form their own judgements on the reliability of the product, as well as to pose the right questions to the evaluation team. Ideally, therefore, the people in charge of the evaluation should have some experience of the practical work of evaluation, having done it in the past.

How much to spend?

It is difficult to decide how much to spend on an evaluation on an *a priori* basis. In general terms for large scale relatively routine programmes the budgets required for evaluation will be a small proportion of the programme resources (normally less than 1%). On the other hand for interventions that are relatively innovative and pilot in character and where evaluation has a strong learning and participatory aspect the costs are likely to be a relatively high proportion of programme (up to 10%).

The most appropriate basis for determining the budget is the nature and scope of the work required. Good evaluation requires inputs from good evaluators and the commitment of those commissioning the work and stakeholders alike.

Budgetary resources should not be a factor limiting the quality of an evaluation. However, there are diminishing returns. At the ex ante stage the time available to inform programme formulation and data availability are likely to be limited. At the interim stage the size of beneficiary surveys and extent of stakeholder consultation will have a major influence on resource requirements. At the ex post stage the quality of monitoring and evaluations that have been undertaken rather than the budget per se is likely to be the main limiting factor.

Who performs the evaluation?

Should an evaluation be conducted by an external team or should it be conducted in house? There are advantages and disadvantages with either solution. External teams will often have greater specialist expertise and may be seen as independent, which can be important for the credibility of the evaluation. In-house evaluators will have greater familiarity with institutional and management requirements and may well have easier access to information and key personnel. They may, however, not be seen as independent and may lack specialist expertise. In part, this relates to decisions about capacity development within public administrations. Some have made a serious long-term

commitment to in-house evaluation capacity located in specialist units. When these are clearly separated from operational management they can overcome concerns about their independence.

There are a number of structural approaches to ensuring the independence of in-house evaluators from programme managers. One approach is to locate the evaluation function in a separate organisational unit or division, for example, in planning or strategy rather than in operations. Another is to ensure that higher levels of management separate from both operations and evaluation are explicitly involved in follow-up of evaluation recommendations and conclusions. This can act as a counter-balance to any tendency to ignore evaluation reports, for example, by holding all parties accountable for follow-up.

However, independence is not only a structural matter. Developing an ethos of independence among in-house evaluators (and supporting a similar ethos among external evaluators) is an important way of ensuring behavioural independence. Furthermore, developing an evaluation culture in the relevant administrative units, one that is self-critical and open to new evidence and to ideas for improvement, can also strengthen the independence of the evaluation function.

There may be different logics appropriate for different stages of the evaluation and programme cycle. It may be preferable to rely more on internal resources for formative evaluation inputs or for ex-ante exercises but depend more on external resources for the ex-post evaluation.

5) Writing the Terms of Reference

The Terms of Reference (ToR) is the document that serves as the basis of a contractual relationship between the commissioner of an evaluation and the team responsible for carrying out the work. Devising the Terms of Reference is a vital step when an evaluation has to be performed by outside consultants. This work is equally important when part of the evaluation is performed in-house. The ToR may concern either the evaluation operation as a whole (when it is entrusted to a single team) or a part of the research work programmed in the evaluation project (in-depth analysis of an evaluative question).

The ToR should be brief (typically 5-10 pages) supplemented if necessary by administrative annexes. A model content for a ToR is listed below and is then elaborated.

Standard layout of the Terms of Reference

- Regulatory Framework
- Scope of the Evaluation
- Main Users and Stakeholders of the Study
- Evaluative and Research Questions
- Available Knowledge
- Main Methods or Techniques to be Used
- Schedule
- Indicative Budget
- Required Qualifications of the Team
- Structure of the proposal
- Submission rules and adjudication criteria

1. Regulatory framework

The legal, contractual and institutional framework for a programme needs to be stated. This would, for example, include regulations of national authorities or the European Union. The ToR should specify who initiated the evaluation project and, where relevant, who was involved in formulating the evaluation brief. Underlying motives and intentions should also be stated. For example: Is the intention a change of policy direction? If so why? Is the intention to modify the implementation procedures? Is the intention to reallocate funds?

2. Scope of the evaluation

We have already discussed the importance of defining the scope of the evaluation. The ToR should clarify the project/programme/policy/theme to be evaluated, the period under consideration, the point of the policy/programme cycle at which the evaluation is set and the geographical area of reference for the study.

3. Main users and stakeholders of the study

We have already noted the importance of evaluation use and users being identified at the earliest stages of planning. It is therefore important to include statements about how the evaluation results will be used in the ToR. If there is to be user-involvement, for example in a Steering Committee, this should also be stated.

4. Evaluative and research questions

We have already noted that different evaluation and research questions can be addressed and different criteria can be employed in formulating evaluation judgements.

It is important to state the evaluation questions but it is also important to limit the number of questions that the evaluation asks. To focus the evaluation on a narrow list of questions that are relevant for the commissioner ensures better quality control.

5. Available knowledge

The ToR should contain a review of the current state of knowledge on the programme and its effects. This will include extracts or references from programming documents, lists of previous analyses and evaluations with relevant extracts, a description of the monitoring system in place, quantified indicators and the various reports and databases available from the services managing the intervention. This inventory is relevant for the evaluation teams to adjust their proposed methods.

6. Main methods or techniques to be used

Each evaluation will have its own particular methods relevant to its scope and content. It is not generally good practice to fully specify methods and approaches but to leave scope for those who propose an evaluation to indicate how they would wish to proceed. The priority is for those who commission the evaluation to specify what they consider to be their requirements in terms of outputs, e.g., answers to key questions. They may or may not specify particular methods consistent with their intentions, for example, the need for a survey of beneficiaries.

The choice is generally made to maintain sufficient flexibility to allow those answering the ToR to differentiate themselves in terms of the relevance and clarity of their methodological proposals. This is especially important in the selection phase because assessing the methodological qualities of the proposals is a crucial step in selecting the right evaluator.

When possible from an administrative point of view, the best way is to determine a budget (see below) and to describe only the main lines of the method in the ToR and then to select the team that

proposes the most promising method. Those selecting the team will then need to have the ability to judge the methodological quality of a tender.

7. Schedule

The evaluation schedule should be established by taking into account various constraints, especially those concerning the decision-making schedule and possible use. It is also necessary to integrate the main deadlines, generated by the procedures of calls for tenders and by the phases of primary data collection. It is advisable to define in the ToR the overall length of the exercise and to leave the first period usually between 10-20% of the duration of the overall evaluation to the detailed planning of the work. This phase should be concluded by an Inception Report in which the design of the activities as well as the detailed schedule will be spelt out. Equally advisable is to define the different outputs of the exercise, and among them, specific reference should be made to the submission of the draft final report allowing enough time for the suggestion of changes and amendments before the end of the study.

8. Indicative budget

It is good practice to suggest an indicative budget and then to leave those competing for an evaluation by open tender to suggest what they would be able to provide for the budget available. This allows value-for-money assessments to be made. It also provides the commissioners of the evaluation with greater control over expenditure. An alternative to this top-down approach is to leave it to proposers to come up with their own estimates based on the tasks they see as necessary. In general, those tendering for an evaluation should be encouraged to breakdown their costs into basic categories, including for example, data collection, report preparation, fieldwork, etc..

9. Required qualifications of the team

The ToR should specify a number of requirements of the evaluation team. This should include: methodological skills required; prior experience of similar evaluation work; knowledge of the regional and institutional context; professional background and disciplinary expertise; and the ability to manage and deliver an evaluation in a timely fashion.

We have already noted the importance of independence in terms of credibility. This can be heightened by entrusting the evaluation to an external team. It is also useful to put in place management arrangements that will support the independence of those evaluators chosen; and request confirmation that there are no conflicts of interest within the potential team. These requirements should be stated in the ToR.

At the same time, how evaluators will be able to have access to key personnel within the programme and its management and to information that they will require for their work, should also be described.

In the field of European Cohesion Policy, a large number of organisations are present in the evaluation market, including local, national or international consultancy firms. The commercial sector accounts for most of the market, although university research centres also make a contribution. Opting for a consultancy firm or a university department can have implications in terms of the approach and therefore the results of the evaluation. Academics have the advantage of being perceived as independent and highly credible owing to their own institutional and professional requirements. Private firms are often more readily available as far as time is concerned and are more concerned with meeting the commissioner's expectations.

The overall choice should depend less on the institutional origins of the evaluation team and more on the required competencies, i.e., their expertise, skill and prior knowledge. Those proposing an evaluation should also be asked to indicate how the different expertise, skills and experience within the team will be integrated and encouraged to work together.

10. Structure of the Proposal

In order to facilitate the adjudication and to provide guidance to the potential applicants, the ToR should specify how the proposal should be structured, possibly indicating the maximum number of pages for each section of the document.

11. Submission rules and adjudication criteria

The tender should specify: the deadline, the modes of transmission (post, fax, e-mail), how long their offer will remain valid, etc.. It should also indicate the criteria according to which the proposals will be judged. The ToR should state for example in percentage points the relative importance that will be given to:

- the quality of the methodological approach;
- the qualifications and previous experience of the team;
- the price.

It is of course important that these criteria are applied systematically once proposals are received.

Implementing and managing evaluations

1) Choosing the right evaluators

The process of selecting evaluators needs to be transparent. Careful drafting of the Terms of Reference (ToR) is the best possible way to attain this goal, together with the use of a formal selection committee. This should include representatives of the people in charge of the evaluation and, when possible, representatives of the potential and actual users of the evaluation. Sometimes, it is useful to include also an independent expert.

Members of the selection committee should reach their own judgements on tenders against the criteria given in the ToR. The judgements should then be combined. The criteria normally include: the quality of the proposed method of approach; the quality and experience of the evaluation team; and the price.

Judging the quality of the proposed method

The suitability of the proposed approach and methods to answer the questions asked in the ToR should be central to the selection of the preferred tender. The selection committee can ensure this by checking each of the tenders against the points in the Box below.

Judgements on the quality of the method proposed are qualitative and not quantitative and need to be made by those with experience. For example, the size of the sample for a survey or a number of case studies may be less important than the quality of the process through which the sample is extracted or the case studies identified.

For each evaluative question

	Question 1	**Question 2**	…
The proposition of the candidate team:			
Does it include the collection of sufficiently relevant information?	++	+	
It is based on rigorous analytical techniques?	-	+	
Is it able to clarify the evaluation criteria in an impartial manner?	+	+/-	
It is likely to produce credible findings?	+	+	
Was the respective importance of the questions well understood?		++	

Judging the qualifications and previous experience of the team

The qualifications and previous experience of the team are always important and especially so if the methods proposed are experimental The capabilities of the team must be matched with the methodology proposed in order to avoid problems occurring whilst the evaluation is being implemented. However there is a danger that this will discriminate against new entrants and therefore make the creation and maintenance of competitive markets more difficult.

A useful way to judge a proposed team is to ask to see previous examples of their work. This can be further supported by asking for references from previous evaluation customers, i.e., named persons who can be consulted.

Finally, it is always good to pay attention not only to the presence in the team of highly qualified personnel, but also to the time that they are prepared to devote to the task. As evaluations are time consuming, the most qualified people will not undertake all the fieldwork themselves. Time allocated by those with experience needs to be sufficient to provide supervision for those working in the field. Evidence of the proposed team having worked together successfully is also relevant.

Assessing the price

Assessing the proposed price for the services is an important aspect of the selection process, but should not be overestimated. Different rules apply in different national contexts. It is not good practice for the price to over-ride the quality criteria.

A second point worth noting is that not only should the total price should be taken into consideration but also the unit cost of the workday for the different categories of personnel employed. For instance, if 80% of the total price is absorbed by junior personnel at, say, a low day rate, then the merits of this can be compared with a situation where 50% of the work is carried out by more qualified/experienced researchers working at twice this daily rate.

In some countries, in order to avoid a race to the bottom the price is judged not in absolute terms but in relation to the average proposed by the teams bidding for the work. In this case, if an offer is exceptionally low, the tenderer could be asked to justify the reasons why such an offer is possible.

2) Managing the evaluation process

Once the evaluation has started there is the temptation for the commissioning authority to keep contact with the evaluation team at arm's length. This view is based on the belief that a hands-off approach will help to secure the independence of the evaluation team.

The independence of the team, in fact, depends on a much more complex set of factors than the reduction of contacts with the client. The best guarantee of independence is the scientific and professional standing of the selected team. The existence of a large market, the emergence of professional and ethical standards and the creation of a community involved in evaluation, are relevant structural dimensions that ultimately support independence.

When managing evaluations commissioners and programme managers need to be aware that there are a number of factors that can undermine the independence of evaluators.

Factors influencing the independence of evaluators

All evaluation requires a measure of independence between the evaluator and the object of evaluation. Normally, increasing the level of independence of the evaluator from the object of evaluation will increase the credibility of the evaluation findings. In all circumstances the possibilities for conflicts of interest need to be minimised and where possible eliminated. Sometimes this is achieved through formal declarations from evaluators and potential evaluators as to the absence of such conflicts.

However, evaluators are rarely fully independent from the objects of evaluation and evaluation is never value free. Evaluators will be subject to a range of influences. Indeed the commitment of the evaluator to the aims of the intervention under consideration may well increase the quality of the evaluation findings and the chances that the results lead to improved socio-economic development.

Several factors influence the independence of the evaluator not all of which are avoidable and sometimes external influences can bring benefits:

- Evaluators tend to be sympathetic to the underlying socio economic development objectives of interventions. They might reside in the territory or have been chosen in part because of their empathy with the target group of the intervention under consideration. Often evaluators are selected because of their knowledge of the relevant theme or policy area and contacts as well as evaluation experience.
- Evaluators generally want to be heard and to have influence. Evaluation is normally both summative and formative and the balance between the two may well shift during the evaluation. If those commissioning evaluation are faced with a new policy choice they may wish the ToR to be changed or may request initial impressions from the evaluator. Early evaluation results might raise serious issues that had not been foreseen and identify the need for radical changes in the intervention proposed or already underway.
- The interpretation of evidence depends upon an understanding of the way in which the world works. The evaluator will have his or her own views on the likely consequences of different types of interventions built upon a combination of direct experience, educational disciplinary background and personal values. In final reports and when justifying a proposed method, these views, experiences and values need to be made explicit.
- The evaluator is normally paid. In most instances those who commission evaluation have responsibility in part or in full for the interventions that are being examined. In some instances evaluation is a requirement of third parties and there may be a temptation for collusion between commissioners and evaluators. Successful evaluation requires a strong measure of trust which can be reinforced by the kinds of standards and codes of ethics for evaluators, described below, and a willingness on behalf of those commissioning the work to listen to the findings of the evaluation and the views of the evaluator.

- Evaluation of socio economic development never takes place within a politically neutral environment. Territories or target groups that have received priority may wish to retain this priority status and the success of previous interventions may be a factor in future access to resources. There may be rivalry between those responsible for different interventions. Those commissioning evaluation are often under pressure to produce programme outputs and evidence of achievements.

The varying roles and circumstances in which evaluation takes place will affect the degree of independence that can be achieved. Where the evaluator mainly provides research inputs and collects evidence, a high degree of independence can be achieved. However, even in these circumstances the choice of questions asked and the method of asking them can condition the independence of findings. Where evaluation is primarily undertaken for scrutiny, inspection or quasi-audit purposes the independence of the evaluator tends to be greater. Where the evaluators work in close cooperation with those preparing the interventions the role of the evaluator has been characterised as that of a critical friend. This often occurs at an ex ante or feasibility stage, though not exclusively. Such evaluations are essentially supportive but willing to point out difficulties and weaknesses in the analyses underpinning prospective interventions. Where the intervention is experimental or being undertaken on a pilot basis, true independence may be difficult to achieve or even be desirable. Here the intervention is relatively small and complex but involves different parties working together for the first time perhaps and the evaluator may be as much an animator and catalyst for consensus as impartial observer. Often evaluation involves a combination of review and case studies where the latter can be used to build arguments. The selection of cases and evidence may constrain impartiality. Whenever the evaluator is in effect an auxiliary feedback loop between actors from different levels of government, there is a particular need for an awareness of professional and ethical standards both among evaluators and partners.

Interaction between commissioner, partners and evaluator

There are a number of reasons why the management of an evaluation requires continuous and meaningful interaction between all the involved partners (including the evaluation team itself).

A first phase during which the team tests and refines the declared justification for the evaluation through consultation with potentially interested parties is usually advisable. An inception or start-up phase will usually specify the methods and workplan in a more detailed way than was possible at the proposal stage. The evaluation team will usually only be able to propose a detailed plan after a first survey of the field and an analysis of the available data. This fundamental part of the evaluation design must be shared and agreed with the commissioner and the other relevant stakeholders.

It is important to allow a certain amount of time between the selection of the evaluation team and the commencement of the work. Particularly when the selection involved a call for tenders, it is unrealistic to expect that the winning team will be able to start working the day after the decision. Given the uncertainties surrounding the choice of the contractor, most applicants will need some weeks in order to plan and assemble the team that will actually do the work. This can be done by delaying the signature of the contractor by allowing an adequate period for the Inception Report.

Role of Inception Report

The Inception Report sets out:

- the main stakeholders identified;
- the most relevant evaluation questions (elaborated and possibly restated);
- the methods to be employed;
- a detailed work plan with the division of labour between the different members of the team;
- the (finalised) schedule for the work, including the various milestones; and

- the intermediary and final outputs.

This document must be discussed and agreed with the Steering Committee in the first meeting following the start of the work. It will represent, for the whole duration of the exercise, the main point of reference of the quality assurance process (see below), as it states in detail what can be expected from the exercise, the points in time at which the different activities will be performed, and the process through which the evaluation reports will be produced.

Interim and Final Reports

In some evaluations, especially those that last longer, there is at least one interim as well as an inception report. This allows for the sharing of first impressions and provides an opportunity to focus the subsequent stages of an evaluation when early findings highlight such a need. This is especially important when evaluations are expected to inform or advise programme management.

For draft final reports, it should be emphasised that in the interests of independence, Steering Committees should concentrate on issues of accuracy and conformance to expectations rather than try to second-guess or influence evaluation conclusions.

The Steering Committee

As we have seen the existence of a Steering Committee or Evaluation Committee is an important part of the process by which evaluations of socio-economic development programmes are managed.

The experience of the Cohesion Policy shows the advantage of involving the most important stakeholders, and in particular the relevant institutional and other key partners whose co-operation is needed to improve the programme. The advantages of an inclusive Steering Committee are shown in Box below.

Advantages of a Steering Committee:

Establishing an evaluation Steering Committee consisting of the different stakeholders in the programme helps ensure:

- **better acceptance of the evaluation by those evaluated, by creating relations of trust;**
- **easier access to information and a better understanding of the facts and events which took place while the programme was underway;**
- **opportunities for process use and learning among stakeholders as a result of their Steering Committee interactions;**
- **interpretations and recommendations which take into account all the important points of view;**
- **the dissemination of conclusions and taking into account of recommendations more rapidly and informally; and**
- **a greater likely-hood that recommendations and conclusions will lead to action and follow-up.**

Generally, the Steering Committee should include four categories of people:

- *The strategic management* of the programme or intervention, i.e. the funding authorities, the policy side of the administration and, where appropriate, the different levels of government;
- *The operational management* of the programme or intervention, i.e. those whose activities are being evaluated, although in order to guard the impartiality of the Steering Committee,

operational management is usually represented by senior managers, a little distant from the front-line day-to-day management. It is an important task of Committee chairpersons to ensure that no members attempt to influence evaluation findings or ignore any body of evidence;

- *The social partners*: i.e., the people representing the main interests affected by the programme. These can include not only trade associations, trade unions and the economic interest associations, but also the institutional and societal bodies in charge of specific, horizontal aspects like the environment, equal opportunities, tourism and consumer protection, etc..
- *The experts*: that is people that have either substantive or methodological knowledge that can be useful for defining the evaluation questions or interpreting the results. The presence of independent experts in the Steering Committee can provide useful inputs to the evaluation team and open up debate towards more general lessons that can and should be drawn from the exercise.

The role of the Steering Committee is to ensure a high quality and useful evaluation. This will involve facilitating the work of the evaluators through, for example, providing access to information and contacts and elaborating evaluation questions and key issues. The Steering Committee should not attempt to influence the evaluators to omit certain evidence or to come to conclusions they would prefer to hear that are not substantiated by the evaluation evidence. The Steering Committee may also oversee the process of communicating the evaluation findings.

3) Managing evaluation communications

Communication is an important part of the evaluation process. It is better to treat the communication task as continuous: an opportunity for dialogue and the accumulation of understanding rather than put all communication efforts into one big dissemination exercise after a final report is delivered. Communication should therefore include:

- Improving awareness of the evaluation underway;
- Providing feedback on interim findings;
- Circulating and managing feedback on draft reports and outputs (e.g. data collection instruments); and
- Communicating evaluation findings and conclusions.

Improving awareness of the evaluation underway

Once the evaluation team has been engaged it is useful to provide information to stakeholders on the timetable and process. The inception period should be used as an opportunity to both explain the planned approach and to canvas opinions on the usefulness of the evaluation questions and the likely success of what is being proposed. In addition to formal information provided to stakeholders perhaps through the Steering Committee, general information to the public and beneficiaries perhaps in the form of press releases or information on websites can also be a useful way to prepare the ground for the evaluation.

Providing feedback on interim findings

The communication of interim findings poses some challenges. On the one hand stakeholders are likely to have a keen interest in early findings particularly if they suggest that the ultimate findings will be critical. At the same time the evaluation team may be hesitant about inferring major conclusions and nervous about the strength of the evidence base for their observations. They may (but should not) also view the production of interim findings as little more than a bureaucratic necessity (interim reports often trigger interim payments). It is best if attention is given in the inception report

to the likely scope and content of interim findings and the method and extent to which they will be circulated.

Circulating and managing feedback on draft reports and findings

Producing the draft final report is often a difficult stage both for evaluators and stakeholders. What has previously been anticipated now becomes real and sometimes threatening or disappointing. Stakeholders, especially those with programme management responsibilities, may be tempted to discredit findings they do not like. Evaluators for their part may construct arguments on limited evidence or be insensitive to the political import of what they present. Producing a final report that is acceptable to the evaluation team and the commissioning authority and respected by stakeholders who have been engaged in the process is a major challenge and requires a good deal of time. The following suggestions may facilitate the process:

- The structure of the report should be agreed as early as possible.
- The Steering Committee should be the main forum for discussion of the draft.
- The contracting authority should avoid the temptation to overly influence the formulation of conclusions and recommendations. Rather they should challenge the evaluation team to justify their conclusions and recommendation on the basis of the evidence presented.
- Sufficient time should be given for written comments.
- The contracting authority should take responsibility for the circulation of the report and compiling feedback.

Communicating evaluation findings

Evaluation is of no consequence unless the findings are communicated. The principal form of communication is a written report. Whilst the appropriateness of the particular means of communication will vary there are a number of good practices:

- The written report should be clearly written and concise. One hundred pages including an executive summary are normally sufficient. Detailed evaluative evidence such as case studies and quantitative analysis should be presented in annexes or made available separately.
- The report should include an executive summary of 5-10 pages written in a style suitable for policy makers.
- The links between the conclusions and the analysis of evidence should be clear.
- The drafting of the report should indicate the basis for the observations made: the evaluation evidence or a combination of evidence and the evaluator's opinion.
- The report should include a description and assessment of the method used that is sufficiently detailed and self-critical to enable the reader to judge the weight of evidence informing the conclusions.
- Use should be made of tables and diagrams where they improve the presentation of findings.
- Reference should be made to good practice examples of interventions to illustrate the arguments being made but evaluation reports should not take the place of good practice guidance. Pressure on evaluators to produce good news stories is often counterproductive: such results are viewed with suspicion by public and policy makers alike.
- The recommendations made should be clear in the follow-up action that is required.

Channels for communicating evaluation findings and reaching users

Those responsible for commissioning and undertaking the evaluation should ensure that the results are communicated and used. Potential users, from policy makers through beneficiaries to the general public, need to be identified and the most appropriate channels of communication selected.

Evaluation reports are normally published, usually on the internet. Written reports should also include more popular outputs for news media to take up. Many programmes produce their own newsletters and these provide another opportunity for dissemination. Oral presentations to the Steering Committee and other stakeholders (e.g., in specialised workshops) are also useful.

4) Managing quality assurance and quality control

Assessing the quality of an evaluation is an integral and fundamental part of the evaluation process. An evaluation that does not meet some minimum quality standards could mislead decision-makers and programme managers.

However, to assess evaluation quality is a complex process. The evaluations performed in the context of socio-economic development programmes and policies are too different from each other to allow the existence of few simple rules that can guarantee the quality across the board. By and large one can say that the quality of the evaluation as a whole is conditional upon the presence of three distinct but interrelated aspects:

- the quality of the planning and design phase, including the commissioning of the evaluation;
- the quality of the implementation of the evaluation itself;
- the quality of the available data.

These aspects are interrelated in the sense that poor performance by the evaluator can very well stem from the poor quality of the data or from the flaws of the planning and design phase. Unfortunately those involved in these three sets of activities are different and often their goals, as well as their quality criteria, are also different. For instance the monitoring system designed for the day to day management of the programme does not necessarily produce the data needed for an evaluation of impacts.

In the first place, quality can be considered a *characteristic of the process* through which the evaluation activities are performed. The assessment of quality could include: the way in which the commissioning authority develops the decision to proceed to an evaluation, defines its scope and the resources available. This can be analysed in order to understand if the procedures followed were appropriate to the allocation of the different responsibilities, if the contribution of the various stakeholders was taken into consideration, etc.. The same goes for the performance of the evaluation. One can focus on the way in which the team, and its interaction with the commissioner and the evaluators, was managed, the checks that were put in place in order to ensure that the data collected were properly treated, etc.. The organisation of the monitoring process can be assessed as well.

In the second place, quality is a *characteristic of the products* of the evaluation process. Thus one could analyse the ToR according to the criteria that we have already spelled out. Of course, one can assess the quality of the intermediate and final evaluation reports to see whether they meet some basic criteria of good professional practice and if the data are sufficient in quantity and reliable enough to warrant sound judgements.

In theory the two aspects the process and the product are linked: a good process should generate a good product.

It is now becoming common to define good practice standards in evaluation. These have been elaborated by international bodies (such as the OECD), National Administrations (for example, the Italian Department for Economics and Finance) or professional associations such as national evaluation societies and associations. Many of these follow on from earlier efforts in the United States and can be traced back to American Evaluation Association (AEA): Guiding Principles for Evaluators (1992) and the Joint Committee on Standards for Educational Evaluation Program Evaluation Standards (1994).

Standards Guidelines and Ethical Codes:

- USA - Program Evaluation Standards (1994) Joint Committee on Standards for Educational Evaluation, Program Evaluation Standards: http://www.jcsee.org/program-evaluation-standards
- Deutsche Gesellschaft fuer Evaluation (DeGEval): Standards fuer Evaluation (2001) : http://www.degeval.de/degeval-standards
- Switzerland: SEVAL Evaluation Standards: http://www.seval.ch/de/documents/seval_Standards_2001_dt.pdf
- The African Evaluation Guidelines 2000: https://www.globalhivmeinfo.org/Gamet/Gamet%20Library/1295_African%20Evaluation%20Guidelines.pdf
- American Evaluation Association (AEA), Guiding Principles for Evaluators : http://www.eval.org/GPTraining/GP%20Training%20Final/gp.principles.pdf
- Australasian Evaluation Society (AES), Guidelines for the Ethical Conduct of Evaluations: http://www.aes.asn.au/about/Documents%20-%20ongoing/guidelines_for_the_ethical_conduct_of_evaluations.pdf
- UK Guidelines for good practice: http://www.evaluation.org.uk/about-us/publications
- PUMA Best Practice Guidelines : http://www.oecd.org/dataoecd/11/56/1902965.pdf
- Italy Treasury Guidelines: http://www.dps.tesoro.it/documentazione/docs/all/Criteri_qualita_sistema_nazionale_valutazione_maggio2002.pdf

This list provides a cross section of some current evaluation standards and codes. Most, in particular those that derive from the AEA Joint Standards such as the German Evaluation Society's (DeGEval) and the African Evaluation Guidelines are directed primarily at the technical conduct of evaluation by evaluators. The Australasian and to some extent the UK Evaluation Society's outputs are more concerned with ethical codes of practice. More recently a further category of guideline has emerged. This is directed more at those who commission evaluations. Examples of this can be found in the OECD (PUMA and DAC guidelines) and in the European Commission.

Although there is not yet consensus about all the components of a quality assurance system for evaluation, we see a shift from a focus largely on quality control, i.e., ways of judging report/ output quality. This shift was endorsed by a number of years ago by study on the use of evaluation by the European Commission.

Quotation from EU research on use of evaluation:

[The study] … tends to support the value of inclusive standards that encompass the interests of commissioners of evaluation, evaluators and citizens. Broader European evaluation standards **(instrumental and ethical) as are being considered by European Evaluation Society and several other European national evaluation societies could complement the move towards standards developed by the European Commission and some National Administrations (The Use of Evaluation in Commission Services October 2002).**

Both quality control and quality assurance criteria are suggested below. Normally the person responsible for managing the evaluation within the commissioning body takes responsibility for applying the quality control criteria. Performance on the quality assurance criteria should be informed by the views of members of the Steering Committee, other stakeholders, the evaluation team and those responsible for managing the evaluation in the commissioning body. The Steering Committee should provide the criteria as early as possible in the evaluation assignment and is normally best placed to make the overall assessment at the completion of the work. For quality control, consultation with external experts or referees can be useful. The application of quality control/content-type criteria and quality assurance/process-type criteria are undertaken for different purposes. Quality control of content offers assurance that the work has been properly conducted and that its conclusions can be relied on. Quality assurance of the evaluation process will contribute more to learning about evaluation management and provide inputs that should improve future evaluation management.

Quality control and quality assurance criteria

Quality Control: Output Criteria	**Quality Assurance: Process Criteria**
Meeting needs as laid out in ToR	Coherent and evaluable objectives
Relevant scope and coverage	Well drawn terms of reference
Defensible design and methods	Sound tender selection process
Reliable data used	Effective dialogue and feedback throughout evaluation process
Sound analysis	Adequate information resources available
Credible results that relate to analysis and data	Good management and co-ordination by evaluation team
Impartial conclusions showing no bias and demonstrating sound judgement	Effective dissemination of reports/outputs to Steering Committee and policy/programme managers
Clear report with executive summaries and annexed supportive data	Effective dissemination to stakeholders

Quality control - output criteria

- Meeting needs

Has the evaluation answered the questions of the ToR satisfactorily and does the report provide additional information that might be essential for the commissioners? In particular:

– Has the way programme or intervention objectives evolved and been analysed?

- Does the report cover the entire programme? If not, is the selection justified as regards the priorities stated by the commissioners in the ToR and subsequently?
- Does the evaluation provide useful feedback for programme managers? Does it include lessons on successes and failures that may be of interest to other programmes, regions or countries?

Relevant scope

Have the essential characteristics of the programme or intervention have been described and have problems and successes in implementation have been properly clarified?

Has the evaluation overlooked other potential or future results, as well as any unexpected yet significant effects and results?

Finally, it is necessary to check whether:

- the limits of the scope, in terms of areas or groups, are defined according to the logic of the intervention;
- the scope includes peripheral areas or non-eligible groups which are nevertheless likely to be affected by the evaluated interventions;
- the evaluation considers the evaluated programme or intervention in isolation or includes its interactions with other European or national programmes.

Defensible design

Methodological choices must be derived from the evaluation questions. The evaluation must, moreover, make the best possible use of existing research and analyses:

- Has relevant knowledge been collected and used wisely?
- Is the method or the combination of methods justified for answering the evaluation questions properly?
- Were the reference situations chosen (counterfactual or other) appropriate for comparisons?

Any evaluation report must include a description of the methodology used and clearly define the sources of data. Similarly, the limits of the combination of methods and the tools used must be clearly described. It is necessary to check whether:

- the methodology is described in enough detail for the quality to be judged;
- the validity of data collected is clearly indicated;
- the available data correspond to the methods used.

Finally, does the evaluation adequately analyse relations of cause and effect for the most essential questions?

Reliable data

Evaluators use secondary data and primary data they have collected for the evaluation. In the latter case, the methods used to collect and process the data are important factors in the reliability and validity of the results:

- Are available sources of information identified and the reliability of this data checked?
- Have data from the monitoring system and previous studies been used optimally?
- Were techniques used to collect data complete and suitable for answering evaluation questions?

Whether the collection of data used quantitative or qualitative techniques or a combination, it is necessary to inquire if:

- the mixture of qualitative and quantitative data is appropriate for valid analysis;
- the "populations" used for data collection have been correctly defined;
- the survey samples or cases studied have been selected in relation to established criteria.

▪ Sound analysis

Quantitative analysis consists of the systematic analysis of data using statistical and other techniques. It has a particular focus on numerical values. Qualitative analysis consists of the systematic comparison and interpretation of information sources with a particular focus on why things happen. In both cases it is necessary to assess whether the methods of analysis used are relevant to the data collected and whether the analysis has been carried out to an appropriate quality. Relations of cause and effect are complex and therefore constitute a particular challenge for evaluation. It is necessary to check:

- whether the relations of cause and effect underlying the programme are explicit and relevant so that the object of analysis can be focused, and
- to what extent the analysis uses suitable techniques.

For this reason, a comparison between beneficiaries and a control group or at least a before-after comparison, is recommended, with an explanation for the choice of approach given.

▪ Credible results

The credibility of results requires that they follow logically and are justified by the analysis of data and interpretations based on clearly presented explanatory hypotheses. The validity of the results means that the balance between internal validity (absence of technical bias in the collection and processing of data) and external validity (representativeness of results) must be justifiable. The need to perform in-depth analyses of a part of the programme poses the problem of extrapolation, from case studies, for the programme as a whole. In this context, it is necessary to check that:

- the interpretative hypotheses and extrapolations are justifiable and the limits of validity have been defined;
- the selection of cases and samples makes it possible to generalise the findings.

▪ Impartial conclusions

Conclusions include suggestions and sometimes recommendations. Whereas results are "technical" and can be analysed without too much risk of impartiality, conclusions and recommendations are issued on the basis of value judgements. The quality of the judgement is thus decisive. To answer the question: Are the conclusions fair, free of personal or partisan considerations and detailed enough to be implemented concretely, it is necessary to check that:

- the elements on which the conclusions are based are clear;
- the conclusions and recommendations are operational and sufficiently explicit to be implemented;
- controversial questions are presented in a fair and balanced way.

The evaluation questions of the ToR must be answered.

- Clear report

Evaluation results can be disseminated and communicated to the stakeholders in writing or verbally. The final report is only one means of diffusion and continual communication of results is desirable. The clarity of the report will depend on the quality of the presentation of results and the limits of the work performed. It is necessary to check that:

- the report was written clearly and is set out logically;
- specialised concepts are used only when necessary and they are clearly defined;
- presentation, tables and graphs enhance the legibility and intelligibility of the report; and
- the limits of the evaluation, in terms of scope, methods and conclusions, are clearly shown.

In many cases only the summary of a report is read so this must be clear and concise. It must present the main conclusions and recommendations in a balanced and impartial manner. It must be easy to read without the need to refer to the rest of the report.

- Quality assurance criteria

The next set of criteria concerns the overall process and context of the evaluation: quality assurance rather than quality control. It will allow those assessing quality both to understand what might account for positive and negative aspects of the evaluation and draw lessons that could be applied in order to improve the quality of future evaluations.

- Coherent and evaluable objectives

The coherence of the objectives: the extent to which they are specific, linked to interventions, not contradictory, etc., has been discussed earlier. It was noted that the use of logic models, programme theory and theory of change approaches are useful ways to clarify programme objectives and the logic of interventions at the early stages of a programme prior to the launch of an evaluation. At this stage we are interested in the outcomes of this earlier process. How far were the evaluators dealing with a coherent programme or intervention in terms of objectives? Were any evaluation difficulties the result of poorly articulated objectives or other problems of evaluability?

- Well drawn terms of reference

Sound terms of reference make for effective evaluations. It is possible at the time they are drafted to judge the adequacy of a ToR. It becomes easier with hindsight to identify what might have usefully been included. This is important for future learning. A poor or incomplete ToR can lead evaluators to deploy their resources inappropriately.

- Sound tender selection process

Was the tender selection process well conducted? This is both a procedural question and a matter of substance. Procedurally an assessment should be made of the systematic application of relevant criteria at selection. Substantively we are interested in whether the right decision was made.

- Effective dialogue and feedback throughout evaluation process

Keeping an evaluation on track, providing feedback and providing a forum for stakeholders to learn through dialogue with each other and with the evaluators is a recognised prerequisite for quality. This is partly a question of the forum created for this purpose, the Steering Committee, but possibly also specific briefing meetings and workshops. The inclusiveness of the membership of such meeting places needs to be assessed: were the right stakeholders involved?

Was good use made of Steering Committee meetings? Were the agendas appropriate? Did stakeholders see these opportunities as productive and enhancing their understandings? Did they ultimately help shape and improve the quality and usefulness of the evaluation?

- Adequate information resources available

Evaluators need information. An assessment needs to be made of the adequacy of information. Monitoring systems should help programme managers and an evaluation usually reveals the extent to which they do. Evaluators will also need to draw on secondary administrative data, gathered often for other purposes by local, regional and national administrations.

Much information in an evaluation is held in the minds of key informants. This is especially so for contextual and qualitative information which is important not only to understand the intervention but also how to interpret more formal data.

To judge the quality of the process and context of the evaluation there needs to be an assessment first of whether information existed and second whether it was made available. Judgements about the availability of information and data to evaluators can provide data about the actual state of partnership and inter-agency cooperation.

- Good management and co-ordination by evaluation team

However well planned and however convincing the work plan and inception report, all evaluations need to be executed properly. They need both to follow plans and be able to adapt to unexpected events that make plans - or aspects of them - redundant. Teams need to be kept together and the different work components need to be co-ordinated and outputs integrated. Relations with commissioners of evaluation, programme managers and a variety of informants, implicated institutions, groups and associations have to be managed. These aspects of management are mainly the responsibility of the evaluation team and its managers. However, some elements are shared with programme managers and those responsible for commissioning the evaluation.

- Effective dissemination of reports/outputs to Steering Committee and policy/programme managers

Report dissemination is another shared responsibility. In part it depends on the ability of the evaluation team to produce high quality and well-drafted outputs. It also requires an awareness of the value and opportunities for dissemination within the evaluation team. There is a big difference between evaluators who confine their feedback to the contractual minimum and those who provide ad hoc feedback when new problems occur or when key issues need to be resolved.

Dissemination also requires sensitivity to the information needs and interests of key stakeholders. Sometimes outputs need to be tailored to meet different interests.

- Effective dissemination to stakeholders

An evaluation process should not be considered complete until a programme of dissemination has taken place. The general requirements for such dissemination should have been signalled in the ToR. Primary responsibility does not rest with evaluators. Programme managers and those who commission evaluations should take responsibility for dissemination, including to the public at large.

The synthetic assessment

A synthetic assessment recapitulates the quality criteria. It is difficult to recommend any particular weighting for the different criteria because their importance varies from one situation to the next. A five point rating scale is used. Thus there are two positive possibilities and two negative possibilities and a mid-point when the balance of judgement is uncertain.

Please assess the evaluation report in terms of your judgements as to how positively or negatively it met each criterion specified below:	**Very positive**				**Very negative**
1. Meeting needs: The evaluation report adequately addresses the requests for information formulated by the commissioners and corresponds to the terms of reference	☐	☐	☐	☐	☐
2. Relevant scope: The rationale of the programme, its outputs, results, impacts, interactions with other policies and unexpected effects have been carefully studied	☐	☐	☐	☐	☐
3. Open process: The interested parties – both the partners of the programme and the other stakeholders – have been involved in the design of the evaluation and in the discussion of the results in order to take into account their different points of view	☐	☐	☐	☐	☐
4. Defensible design: The design of the evaluation was appropriate and adequate for obtaining the results (within their limits of validity) needed to answer the main evaluative questions	☐	☐	☐	☐	☐
5. Reliable data: The primary and secondary data collected or selected are suitable and reliable in terms of the expected use	☐	☐	☐	☐	☐
6. Sound analysis: Quantitative and qualitative data were analysed in accordance with established conventions, and in ways appropriate to answer the evaluation questions correctly	☐	☐	☐	☐	☐
7. Credible results: The results are logical and justified by the analysis of data and by suitable interpretations and hypotheses	☐	☐	☐	☐	☐
8. Impartial conclusions: The conclusions are justified and unbiased	☐	☐	☐	☐	☐
9. Clear report: The report describes the context and goal, as well as the organisation and results of the programme in such a way that the information provided is easily understood	☐	☐	☐	☐	☐
10. Useful recommendations: The report provides recommendations that are detailed enough to be implemented	☐	☐	☐	☐	☐
In view of the contextual constraints bearing on the evaluation, the evaluation report is considered to be:	☐	☐	☐	☐	☐

The use of evaluation and knowledge management

Undertaking evaluation and ensuring its quality is only worthwhile if the activity leads to some use of the evaluation findings and contributes to improved knowledge amongst those best able to take advantage from it. There are at least three different ways in which evaluation is used.

- Evaluations may be used directly in an instrumental manner when the results, findings, conclusions and recommendations are taken up. In practice this is unusual and where it does occur it tends to take place only partially.
- More often, several evaluations or evaluations combined with other evidence and opinion are used cumulatively to inform debates and influence decision-making. Evaluation thus stimulates the process of debate, challenge and counter challenge to evidence and its interpretation.
- Even where evaluation results are not used the process of evaluation initiation and reflection can be useful by offering opportunities to exchange information and clarify thinking.

The extent of use of evaluation and its impact is influenced by a number of factors:

- The organisational arrangements for dissemination. The time and resources available for dissemination and the degree to which the process is championed by those responsible for the work influences the understanding, communication and use of the findings.
- The quality of the evaluation. Where evaluation standards are high the results cannot be easily dismissed.
- The involvement of stakeholders in the stages of the evaluation cycle alongside evaluators and administrators. This is essential to build up evaluation use.
- The involvement of senior managers. This helps ensure that policy and resource allocation as well as practice are influenced by evaluation findings.
- The application of a system of systematic follow up of the conclusions of evaluation. This process both draws attention to where the findings have been and have not been used and reduces the tendency to relearn the same lesson.
- The institutional arrangements for conducting evaluation. There are no perfect models. Evaluation findings are likely to be of use to decision makers, those involved in the planning and design of interventions, and those involved operationally. The tendency towards the organizational separation of evaluation, operational and policy functions may lead to the improved independence and quality of evaluation. Policy and operational concerns can for example over emphasize what can be achieved through evaluation. On the other hand the separation but may be less helpful if it leads to an overemphasis on evaluation management and limits the use of the evaluation.

It is reasonable to conclude that the creation of an evaluation culture is essential for organisational learning. Key components of an evaluation culture over and above the generation of quality evaluations include: a presumption that interventions should be designed and implemented in a manner that facilitates subsequent evaluation; an appreciation of the range of purposes of evaluation; a recognition of the limits of evaluation, the scope for interpretation and the need to combine quantitative and qualitative evidence; and a recognition of the needs of different users of evaluation.

Golden rules

1. Evaluation competence should be brought in early by planners. This can help to clarify objectives. This activity, although employing evaluation competence, is quite separate from mainstream evaluation activities. It needs to occur at the design and planning stage. However, this can make subsequent evaluation easier and more successful. Evaluability assessment and preparing an analysis of programme theory can be used.
2. A similar evaluability assessment should be undertaken by evaluators when they begin their work. This may overlap or repeat what has already taken place at planning stage. However, the purpose is different. It is to ensure that a feasible evaluation plan is produced and to clarify how evaluation outputs will be used. This is consistent with a general expectation of evaluators that they should be concerned with how their results, conclusions and recommendations are used from the earliest possible stage of their work.
3. Stakeholders, programme managers and policy makers, potential beneficiaries and partners should be involved in the evaluation from the earliest stages, where practicable. This will ensure that the evaluation design and plan will include their priorities. It will also ensure that they feel some sense of ownership of the outputs of the evaluation and are more likely to find them useful and use them. It may, however, be necessary to be selective in deciding on which voices finally determine the evaluation agenda, to retain focus and ensure the evaluation is manageable. Overarching priorities should be shaped by the intentions and logic of the programme or intervention whilst remaining open to unintended consequences especially for intended beneficiaries.
4. Evaluations need to be actively but sensitively managed. This will ensure that commissioners are aware of choices to be made along the way and that evaluators receive sufficient support, access to information and briefing as to changes in policy and context. Those responsible for commissioning an evaluation and programme managers are the most suitable people to manage the evaluation because they are aware of its background and rationale.
5. It is usual to derive criteria for an evaluation, i.e., judgements as to the basis for positive and negative assessments of progress, from the objectives of a programme. It is also important to include a wider set of criteria that derive from social needs. For example, is this programme useful for those whom it is intended? Is the programme consistent with other policy initiatives? Maintaining this broader perspective ensures that for part of their work evaluators can stand outside the logic of the programme and take a critical perspective on what it is trying to achieve and how it does it.
6. The importance of evaluation questions in an evaluation design cannot be overstated. The temptation otherwise is to gather large quantities of data and produce sometimes technically sophisticated indicators which make little contribution to practice or policy. There is, of course, a problem formulating the evaluation questions in a way that they are likely to be able to be answered. While this is a technical question and this chapter has offered suggestions about how to formulate questions appropriately, there is here also the overarching concern for use. Ask questions that someone will find useful. However, use should not itself be defined too narrowly. We are talking here not just about the instrumental use of evaluation by managers. We are also talking of uses that citizens and civil society groups may make of evaluation in support of democratic processes and accountability.
7. We have specified in some detail the content and form of ideal Terms of Reference for an evaluation. This is part of the general question of design and the choices that can be made at the design stage which can influence the quality and direction of an entire evaluation. It should be recognised that defining scope, clarifying the users of the evaluation and deciding the skills required for an evaluation team, are among the most important decisions made during the course of an evaluation.

8. It used to be common to regard the use of evaluation as being confined to acting on recommendations and final reports. It is now understood that evaluation use can be supported and occurs throughout an evaluation. Process use should involve stakeholders in evaluation thinking from the beginning. There are evaluations where the conclusions and recommendations are rejected but stakeholders, especially those involved in the steering committee find the evaluation useful. It can help them to clarify their own thinking and understanding and spark off innovative ideas for improvements. Promoting dialogue during the course of an evaluation is likely to ensure that when stakeholders receive reports they will be better prepared and receptive.
9. It is often easier for programme managers and those who commission an evaluation to confine judgements of evaluation quality to the outputs in reports of the evaluation itself. However, this quality control process provides few opportunities for learning and improvement in the way the evaluation itself is managed. A quality assurance perspective provides a context in which to explain the strengths and weaknesses of evaluations. It offers an opportunity for those who commission evaluations to learn how to improve in future.
10. Consideration should be given at an early stage to how evaluation findings will be used. Some use will stem directly from the findings and recommendations of the work. Evaluations can also be helpfully combined with other evidence to inform debates. The process of evaluation can bring benefits in terms of structuring inquiry and institutional reflection. Close attention to the factors that influence the use of evaluation work will maximise its contribution.

Chapter 3: Developing capacity for socio-economic evaluations

Evaluation capacity is part of institutional capacity

We have already painted an ambitious picture of what is expected of evaluation, how it should be organised and what it can deliver. In the previous Chapter, for example, we provided guidance to administrations that commission or use evaluations as to how they might design and implement an evaluation. This assumed that requisite institutional and administrative capacities were available. For our purposes institutional capacity is defined broadly to include legal rules (e.g., regulating employment and procurement) normative assumptions (e.g., about equal opportunities or open competition), governance (e.g., democratic accountability and divisions of responsibility between tiers of government and civil society) as well as administrative and organisational arrangements (e.g., how ministries are structured and resourced). In this Chapter we are concerned with how to create such capacity so as to make practicable the ambitions for evaluation that can contribute to socio-economic development.

Capacity cannot be created overnight nor is it created without costs. However, the potential benefits of evaluation are large enough to justify the investment and the recurrent costs needed to continuously innovate both in evaluation processes and products. It takes time to develop capacity and the necessary systems need longer-term nurturing to deliver sustainable benefits

Developing evaluation capacity has to be a shared concern of the wider policy community. Those responsible for policies and programmes must first be convinced of the need for evaluation. At the same time, even though their support is essential, they must not be allowed to capture the process. To begin with developing evaluation capacity includes those who manage and commission evaluations, those who have an interest in evaluation results at a policy and programme level as well as those who undertake evaluations. However, evaluation requires an institutional framework and has to be embedded in institutional arrangements. The capacity of public institutions to conduct evaluations is part of the wider requirements that the State must meet to address contemporary economic and social demands. Where evaluation capacity has been most developed is often in the very sectors that have conceived of it as an integral part of a much wider programme of public sector innovation and modernisation.

The need to build institutional and administrative capacity is a sometimes implicit but increasingly explicit transversal goal of socio-economic development policy. This goal in turn stems from two imperatives:

1. To overcome the inefficiencies of traditional public administrations by shifting towards a public management approach that draws on best management practice not only from the public sector but also taking on board lessons from the most successful private and non-profit organisations. Many of the assumed limitations to efficiency of public sector bodies are encapsulated in the so called New Public Management (NPM) movement which advocates reform, modernisation, decentralisation and privatisation among other solutions in the public sector. NPM aims to improve results, delivery and value for money.
2. To overcome the perceived distance of public bodies from society as a whole, and therefore open up policy making to the contribution of external stakeholders, civil society representatives and citizens (the global drive towards democratic governance).

The diffusion of evaluation can contribute to both of these imperatives.

What is evaluation capacity?

Evaluation capacity is multi-faceted and needs to be located at many different levels that reinforce each other:

- An Individual level consisting of necessary skills and competencies;
- An Organizational level of management arrangements and structures;
- An Inter-organizational level that bridges public and private bodies through networks, procedures and partnerships; and
- A Societal level that embeds evaluative thinking in civil society including professional organizations - as well as in the public sector.

Cutting across all these levels are institutional factors, regulations, laws, resources, norms, etc., that necessarily underpin organizational arrangements and mechanisms.

In an ideal situation one would expect that many of the following arrangements would be found:

At an individual level:

- There are people throughout government who have experience and skill in evaluation and this is reinforced and renewed by appropriate recruitment, training and professional development.
- Training courses and diplomas are available, delivered by universities, private training providers and professional bodies, open to evaluation practitioners and commissioners.

At an organisational level:

- Evaluation is routinely undertaken at each stage of policy and programming: from planning through to implementation and follow-up.
- Evaluation findings are integrated into decision-making when deciding what policy options to choose, how best to implement and when assessing what has been effective.
- Managers look to evaluation as one important input that will help them improve performance and manage for results.
- A regular flow of evaluations are commissioned that cover the broad spectrum of policies and programmes.
- There are follow-up procedures to ensure that evaluation recommendations are taken seriously and, where feasible, acted upon.
- There are procedures to accumulate evaluation findings and lessons learned so that programme managers and policy makers have an accessible evidence base and an organisational memory.

At an inter-organisational level:

- There is coordination through a network of dedicated evaluation units or functions - to ensure sufficient consistency in the way evaluations are commissioned, managed and executed across government and ultimately across the public sector.
- There are requirements that evaluations take place embodied in legislation, articulated policies and regulatory activity (e.g., audit or parliamentary reviews).
- There is a well-defined market with clear rules so that potential evaluation providers can organise themselves to respond to tenders, complete evaluation assignments on time, develop expertise and understand the priorities of commissioners of evaluation.
- There is a culture of evaluation that values professional standards, independence, learning from experience and evidence based policy.

At a societal level:

- Open and systematic dialogue is maintained between policy makers and evaluation specialists so that priorities for evaluation can be identified and scrutinised.
- There is an evaluation community of practice whose members may work for universities or consultancies or be independent evaluation practitioners but still share a common ethos and standards.
- Evaluation associations exist that bring together those who commission, provide and use evaluations and reinforce a culture of evaluation, disseminate good practice and safeguard the independence of evaluation functions and practitioners.
- There is an awareness of evaluation activity and outputs and a dissemination of reports and findings such that evaluations will be routinely used by various stakeholders (in parliaments, civil society, etc.) to support democratic accountability and transparency.

Developing evaluation capacity

In some countries evaluation capacity evolves organically over a long period of time. This was the case in the UK, Netherlands and Sweden. In most countries, however, there is a need for a strategy if evaluation capacity is to become useful within a defined time-scale. This was the case in most Member States joining the EU in 2004 and 2007. There are many key decisions to be made when starting to develop evaluation capacity in a strategic way. Among the most important are:

- Architecture: locating and structuring evaluation functions and their coordination;
- Strengthening evaluation demand: ensuring that there is an effective and well managed demand for evaluations;
- Strengthening evaluation supply: ensuring that the skills and competencies are in place with appropriate organisational support;
- Institutionalising evaluations: building in evaluation to policy making systems and across the broader policy system.

The "Architecture" of evaluation systems

It is important that those responsible for developing evaluation capacity are aware that they are developing a system in which the various elements will have to interact and reinforce each other. These elements will include, for example, responsible units or divisions in and across departments, procedures, human resource policies, information technology systems, etc.. In these circumstances a strategy needs to be somebody's responsibility. Usually ownership at least in the early stages of developing evaluation capacity is located in a dedicated unit within central government. This might be in a planning department, a national Treasury, a coordinating department, the prime minister's office or a department responsible for EU Cohesion Policy.

Two main architectures have been most common at the start-up stage of developing evaluation capacity:

- A central evaluation unit or function which from the beginning is also associated with a network of other ministry based or sectoral or programme based units. The central unit provides an umbrella organisation within which basic procedures, guidelines, commissioning and public procurement protocols are agreed. Once a system is established the central unit continues to orchestrate but other units increasingly have an equal voice and represent their own priorities within a network.
- A central unit that is located in one department at the outset. It is intended once such a unit is established to diffuse the model that has been piloted and proven itself. Such diffusion might

be formalised as when it is decided centrally that all ministries will in future have evaluation units. Alternatively a more informal process might be chosen: diffusing good practice and encouraging replication.

There are advantages and risks associated with both approaches. A central evaluation unit or function has to strike a fine balance between ensuring that a coherent system evolves - whilst not centralising too much. A balance has to be maintained between adapting systems to local circumstances and ensuring consistency. Centralisation in one coordinating ministry can risk undermining ownership and support for evaluation elsewhere in government, the public sector or among other stakeholders. Too much de-centralisation on the other hand - for example having different coordination arrangements for different programmes, policies or territories, risks incoherence.

Once an evaluation system is established there are further options for location and structure. For example, some ministries or public agencies may concentrate their evaluative efforts in their own central unit, whilst others may rely on internal networks with one or two individuals in a central node and others dispersed across different policy departments, units or divisions.

If evaluation is to become a valued function within the public sector it must itself mirror the architecture of governance within the country concerned. For example, if regional or federal tiers of government are important, then evaluation must also be embedded in these sub-national governance structures. Although this may not be where evaluation is first located, new evaluation systems need to be designed from the beginning with a vision of how they will be adapted more generally.

Strengthening Evaluation Demand

1) Governments as a source of demand for evaluation

The demand for evaluation, the first push that kicks-off the process of capacity development, can come from internal or external sources. In many European countries Cohesion Policy regulations provided the first impetus so the demand was at first external. This was true for southern European countries such as Greece, Italy and Portugal and most of the EU12. This external demand is in part an outsourcing of the responsibilities of European institutions that are obliged to be accountable. The European Commission has obligations to the Member States, the European Council, European Parliament and the European Court of Auditors to ensure that the monies made available for socio-economic development are spent wisely and that the policy goals of better economic, social and territorial cohesion are achieved. Because the same governments that provide the financial resources are the recipients of the money, it is only natural that the evaluation is entrusted at least partly to them instead of only being carried out at a European level. Similar external demands for evaluation are common in international development. Donors, whether individual countries or multilateral agencies such as the EU itself, international banks (World Bank, European Bank for Reconstruction and Development, etc.) or even UN agencies, all require that beneficiaries of funds demonstrate that they have used monies wisely.

Alongside the accountability motive, external actors who initiate a demand for evaluation are also interested in effectiveness. However this motivation is even stronger when the impetus for evaluation is internal.

Many of those who have embarked on evaluation capacity development for domestic reasons have done so as part of a commitment to an effective, efficient, transparent and accountable public sector. Evaluation in Canada, the Netherlands, Sweden and the UK did not begin with external demands. Rather evaluation demand was driven by successive waves of internal reforms of the public sector aimed to improve its management and effectiveness and new notions of governance. This was also true for early stages of public sector reform in Australia, which was mainly driven by a results-based philosophy also reflected in the emphases given to evaluation and its utilisation.

Even when the driving force behind the establishment of evaluation comes from within national borders, a certain degree of external scrutiny is likely. This may take several forms:

- Parliaments, at national, regional or local level, which seek to make government responsible for the efficient and effective implementation of the decisions they have taken;
- Courts of Auditors and similar bodies, wishing to expand their role from the verification of legality and financial control to performance audit, including notions of effectiveness, efficiency and value for money;
- Central government that finances the activities or the investment of sub-national authorities and make it a condition that the latter introduce evaluation in their practices and open themselves up to some form of external scrutiny.

If the impetus for evaluation is entirely external (or even externalised by government departments or agencies to national scrutiny bodies) the foundation for building evaluation capacity is likely to be weak. There has to be a recognition that evaluation is more than an obligation to someone else for demand to be taken seriously. This principle is central to debates about the need for regulatory or legal force behind evaluation. There are those who argue that without a clear legal demand backed up by appropriate compliance mechanisms evaluation will not be taken seriously. On the other hand there is also a strong argument that regulation without cultural, managerial and administrative change will not be effective. There will be token compliance rather than wholehearted adoption of new practices.

There are a number of reasons why systems based exclusively on rules and regulations may not work:

- They depend for their implementation on some kind of sanction. However the ability or willingness of the relevant political authorities whether European Commission, Member States' central government or sub-national authorities, to exercise such sanction is often limited. These limitations may derive as much from political considerations as from legal or administrative capacity.
- With limited likelihood of sanction, evaluation requirements are taken less seriously and evaluation processes and documents tend to be disregarded by the same bodies that prescribed them.
- Managers pre-occupied with their day-to-day responsibilities will minimize their time commitment to evaluation if it is not seen as directly relevant and useful to them. It is unlikely that evaluations designed to meet the needs of external parties will be relevant and useful.

The implications of this latter point are profound. Strengthening demand for evaluation requires the mobilization of a broad coalition of internal stakeholders including managers and policy makers. It cannot simply be created by one or two committed evaluation champions or by deciding to establish a small central unit. Once such a coalition has been established then further procedures and policies can be developed that will reinforce initial demand. Many of the activities identified in Chapter 2, concerned with the commissioning and management of evaluation become relevant. However the strongest boost will come from embedding evaluation in the routine practices of public agencies or government ministries.

2) Civil society as a source of demand for evaluation

A sustainable evaluation system cannot be confined to governments or even the public sector. Civil society can also demand evaluation. Strengthening evaluation is often thought of as a top-down national or even supra-national process. Accountability and learning take place among policy makers and other actors at lower levels are expected to comply with the requirements of the policy community. In terms of contemporary policy thinking this top-down view is anachronistic. Participation, active involvement, responsibility and transparency all assume that civil society actors

are engaged with public policy making and delivery. Capacity development has to include civil-society actors if it is to be true to the logic and values of public sector reform. This can be achieved through public consultations, seeking inputs into priorities for evaluation plans and, as suggested previously, ensuring a broad representation of interests on steering committees and advisory groups.

Supply side

1) Starting Conditions and Traditions

However strong the demand for evaluation, the response to that demand will depend on pre-existing resources, skills and institutions. To a great extent it will depend on the skills, resources and institutions outside the administrative and public management system. There is not a single way to improve evaluation supply. Three approaches are common:

- To build up relationships with educational institutions, in particular, universities;
- To develop and strengthen an independent community of consultants; and
- To support the development of a professional evaluation community.

2) Links with the education sector

Educational institutions and universities in particular are important for the supply side of evaluation for two reasons. First, universities are in many countries the natural place to locate training and education programmes necessary to develop evaluation skills and knowledge. Second many working in educational institutes already have much of the knowledge of social and economic research, data gathering and analysis and report writing that are needed by evaluators. Educational and research perspectives need to be supplemented with other methods, theories and practices to become useful in evaluation, however.

Establishing a working relationship with universities and other educational institutions can serve to stimulate the offer of formal training programmes, continuing education and the creation of centres of excellence linked to research and able to disseminate knowledge. Courses can be directed at graduate students in social sciences and economics, specialists in policy and regional studies and practitioners in socio-economic development. Depending on the target group, it may be best to create new specialised courses or to internalise evaluation modules, e.g., theories and methods of socio-economic evaluation into the formal curricula of economists, engineers, sociologists, planners, etc..

It is important when planning education and training programmes to recognise practitioner aspects of evaluation. Working in a client and contract environment, assessing what will be most useful for programme managers, planning and team management and dealing with multiple or even conflicting interests tend to be more common in an evaluation rather than in an academic environment. Practitioner skills are needed as well as academic knowledge. This can be achieved by including within curricula workshops, training seminars, guest lectures from practitioners, work placements with existing evaluation consultants, study visits to administrative bodies, etc..

In some countries there is a tradition of universities and institutes actively engaging in policy research. In these countries there will be a pool of skills that can be tapped into to increase the supply of evaluators in the relatively short-term. However, even in such countries, this can require cultural change among those who are used to a more academic pace and style of research. In these cases the increase in evaluation supply will take time to mobilize, beginning perhaps with awareness-raising seminars and then specific small scale assignments that will allow potential evaluators to test out new ways of working in a policy environment.

3) Independent Consultants: encouraging new suppliers

In many fields of evaluation there is an international market for evaluation. Large international consultancies are used to bidding for work in a number of countries, not just where their headquarters might be located. It is therefore common for early entrants to new evaluation markets to come from among these international consultancies. They have the potential to import a wealth of international and policy specific experience as well as experience of working with policy makers and programme managers. However, it is important that there should be a home grown evaluation consultancy sector, locally based consultants who are familiar with national institutions, policies and languages. Of course international consultants will often also recruit in national counterparts, thus encouraging international exchange even within a single evaluation team. Such joint working can be an effective method of skill transfer as well as knowledge dissemination. This is widely recognised in different evaluation sectors. For example in international development programmes it is often a requirement that evaluation teams include experts from the northern and southern hemispheres.

There are various approaches to encouraging the emergence of national consultancy suppliers. These include:

- Commissioners of evaluation insisting on consortia or partnership bids that always include some local consultants;
- Scaling evaluation contracts in ways that relatively small low-risk evaluations can be undertaken by national new entrants to the evaluation market;
- Ensuring that procurement and financial requirements associated with bidding for evaluations are not too restrictive (responses required within short time periods, short periods also for bank guarantees and years of audited accounts);
- Emphasising in ToR technical and know-how criteria rather than complex administrative procedures with which less experienced consultants may not be familiar;
- Holding briefing meetings with potential consultants to answer questions and encourage bids in a competitive environment; and
- Support for networking among relatively isolated evaluation consultants so as to encourage team-building, consortia formation and other professional networks.

4) A professional evaluation community

It is in the interest of those who commission and undertake evaluations that they conduct themselves within a framework that encourages high standards. This is what is understood when evaluation is described as a profession. Professional standards are essential for evaluators to be regarded as credible and for evaluators to be granted necessary independence by those who commission and use evaluations. This is especially important in the face of the different and sometimes very powerful interests involved in programme and policy evaluation. Many evaluators have experienced attempts to interfere with how evaluations are conducted. This can take various forms:

- Unbalanced ToR;
- Undue influence on an evaluation process by those who are responsible for the programme being evaluated;
- Attempts to pre-select an evaluation team to include those who will come up with the right answer;
- Pressure to reshape findings and conclusions to be less critical and more acceptable to the main stakeholders; and
- The provision of limited or selected data.

In reality it is in no one's interests for this to happen. For a modern administration the benefits of impartial and trustworthy evaluations far outweigh the apparent benefits of unchallenging and safe reports. However, it is only with the development of an evaluation culture that this becomes apparent.

Professional standards in evaluation are usually assumed to include a commitment among both evaluators and commissioners to:

- Using the best/most appropriate methods;
- Self development and upgrading of skills;
- Taking seriously the interests of stakeholders and policy customers, including programme managers;
- An independent stance that is impartial and true to the evidence collected; and
- Ethical behaviour towards various parties: commissioners, those who provide information and data, potential beneficiaries of programmes and policies.

There are various sources of professional strength and independence. They include:

- An evaluation culture that reinforces professional norms of behaviour;
- Respect for independence by those who commission evaluations;
- Ethical codes on which there is a consensus and which are therefore widely recognised and disseminated;
- Independent institutions within which evaluators work which can lend their judgements greater weight and allow them to resist external pressures;
- High quality education and training and professional development;
- Professional associations and societies that bring together all those involved in the evaluation community where experience can be shared and practical problems discussed; and
- The development and application of quality standards for practitioners and commissioners of evaluation - such as those suggested in the previous Chapter.

In practice many of these desirable attributes of a professional evaluation community can be encouraged under the umbrella of a national evaluation society or associations. These typically bring together people from different institutions (academia, public administration, research centres and consultancies) with different disciplinary expertise (sociologists, economists, planners, political scientists, etc.) and different fields of specialisation (social services, education, research and development and of course socio-economic development). Such societies have become widespread across Europe in recent years. Such societies provide an opportunity for cross fertilisation of ideas, the discussion of shared problems outside of the setting of particular evaluation contracts, the dissemination of good practice and the emergence of common standards and codes of ethics.

List of Evaluation Societies and Associations:

- **European Evaluation Society:** http://www.europeanevaluation.org/
- **Danish Evaluation Society:** http://www.danskevalueringsselskab.dk/
- **Finnish Evaluation Society:**

  http://www.eukn.org/E_library/Economy_Knowledge_Employment/Training/Training/Finnish_Evaluation_Society_FES
- **French Evaluation Society:** http://www.sfe-asso.fr/
- **German Evaluation Society:** http://www.degeval.de/
- **Italian Evaluation Society:** http://www.valutazioneitaliana.it/

- **Spanish Evaluation Society: http://www.sociedadevaluacion.org/website/**
- **UK Evaluation Society: http://www.evaluation.org.uk/**

Institutionalising evaluation

Much of this Chapter has been concerned with the initiation of evaluation capacity and with arguments about why this is important both for socio-economic development and for good public management. Institutionalisation is more concerned to embed and deepen evaluation within public sector institutions. This usually involves:

- extending evaluation more widely within the public sector, and
- greater integration of evaluation processes and utilization into policy making and programme management.

It is likely that the introduction of evaluation will first be concentrated in one part of a government. Extending evaluation spreads evaluation practices to other ministries or programmes and to other institutions. In mature evaluation systems evaluations are used by national audit bodies and parliaments as well as by ministries, policy makers and programme managers.

Integrating evaluation more closely into policy making and programme management is likely to involve:

- The joint planning of policy and evaluation through multi-annual cycles of consultation resulting in multi-annual or departmental plans that include evaluation;
- Systematic follow-up of evaluation findings and recommendations such that it becomes difficult for middle managers to ignore evaluations as they need at least to say why they are not following recommendations;
- The creation of forums where policy makers and managers and evaluation managers meet to discuss evaluation findings and lessons learned; and
- The extension of evaluation throughout the policy cycle from option identification to planning, programming, implementation and reprogramming to the design of the next phase of policy.

Creating an evaluation culture suggests that a well developed evaluation system is more than just a structural arrangement. For example:

- There is a commitment to learning lessons and improvement;
- There is avoidance of a blame-culture which discourages learning;
- Policy makers are committed to evidence based polices in the broadest sense;
- There is a commitment to excellence, high standards and continuous improvement;
- Evaluation is used as one element in a general move towards transparency and multiple accountabilities to citizens and communities as well as to Ministers and parliaments; and
- The government and public sector is committed to continuous adaptation to becoming a learning organisation.

Canada is widely seen as a country which has been successful in institutionalising evaluation into its public management and policy making systems at federal (i.e., national) level. Overall responsibility for evaluation coordination is vested in the Treasury Board for Canada which must provide central direction for evaluation in the Government of Canada and, to that end, should[2]:

[2] http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12309

1. establish a Centre of Excellence to provide leadership, guidance and support to the practice of evaluation;
2. use evaluation results where appropriate in decision-making at the centre;
3. set standards; and
4. monitor evaluation capacity in the government.

Distinctive features of Canadian evaluation policy are that it:

- Relies heavily on the devolution of responsibilities to develop and use evaluations to government departments (ministries) and to managers within these departments,
- Focuses on results based management as a component in public management this, not as a backward looking accountability focus, although accountability is also emphasised

Assessing Evaluation Capacity Building:

The authors of an international comparison of evaluation capacity used the following nine criteria to assess the extent of evaluation capacity:

1. **Evaluation takes place in many policy domains**
2. **There is a supply of domestic evaluators in different disciplines**
3. **There is a national discourse about evaluation**
4. **There are professional organisations of evaluators**
5. **There is a degree of institutionalisation of evaluation in government**
6. **There is a degree of institutionalisation in Parliament**
7. **There are many institutions of evaluators conducting evaluations within each policy domain**
8. **There is an evaluation function within the national Audit Institution**
9. **There is a balance between evaluations that assess outcomes and those that assess outputs and processes**

International Atlas of Evaluation, Editors Furubo, Rist and Sandahl, Transaction Publishers, 2002

Institutionalisation of evaluation is a continuous process which over time has to be integrated with other planning and assessment tools (e.g., impact assessments, project planning techniques) and other channels for identifying and disseminating and implementing good practice (e.g., public consultation, decentralisation to local stakeholders). At the same time potential barriers to institutionalising evaluation cannot be ignored:

- Evaluation is stronger when it is seen as an integral part of institutional development and public sector reform. If these broader organisational and cultural changes are not pursued in parallel it will be more difficult to institutionalise evaluation.
- A coherent institutionalisation process requires both financial and human resources. Institutionalisation can be undermined if there is insufficient investment in the skills and management systems for specialist evaluation units.
- High-level political commitment is also important in the evaluation institutionalisation process. This is especially so when such a process inevitably takes time and needs to be built-up in stages. Changes in direction when officials or ministers change can be a barrier to the successful building of evaluation capacity

As in other Chapters, key lessons are summarised below in terms of golden rules for those seeking to strengthen evaluation capacity.

Golden Rules

1. Evaluation is most useful and used when it is embedded as a function in supportive administrative and institutional systems that seek also to include civil society stakeholders. This is what is meant by evaluation capacity.
2. Evaluation capacity can develop organically over time but in most national and international settings it needs to be planned for, consciously introduced and appropriately resourced.
3. The principle of independence and impartiality is an essential principle to build in to evaluation functions given the vested interests that are likely to seek to influence the conduct of evaluations and their conclusions.
4. Evaluation capability consists of elements at individual, organisational, inter-organisational and societal levels that should reinforce each other. Evaluation capacity should be designed as a system.
5. Evaluative activity needs to be coordinated across administrations to ensure coherence and provide a basis for skills development and professionalisation. At the same time there must be sufficient flexibility in the procedures and approach to adapt to the specificities of policies, programmes and territories.
6. Evaluation capacity may begin by focusing on particular programmes, e.g., those funded by the EU or those taking place at a particular tier of government (e.g., central government departments) but should be designed so that evaluative activity can be extended for example to include nationally funded programmes or regional as well as central government activity.
7. Evaluation should be configured to meet real decision-making and management needs that will improve the effectiveness and efficiency of public action and yield concrete benefits for citizens and stakeholders. Whilst evaluation will inevitably be useful to fulfil external accountability needs, it should never be primarily regarded as an external obligation.
8. The inclusion of civil society and the encouragement of participation in evaluation, whether through consultation or direct representation will encourage different interests (e.g., community representatives, professional groups and the private sector) to demand and use evaluation, thus strengthening accountability, increasing transparency and reinforcing social capital.
9. Evaluation capacity must include attention to the supply side, promoting a skilled and professional community of practitioners and the institutions, networks and markets that will ensure that evaluation assignments are bid for and delivered to a high standard.
10. Among the key underpinnings of a professional evaluation community are education and training for evaluators and quality standards and ethical codes agreed and shared among the commissioners and practitioners of evaluation including academics and consultants. These supply-side qualities are often best developed through national evaluation societies with open memberships and a commitment to high standards.

Chapter 4: Choosing methods and techniques

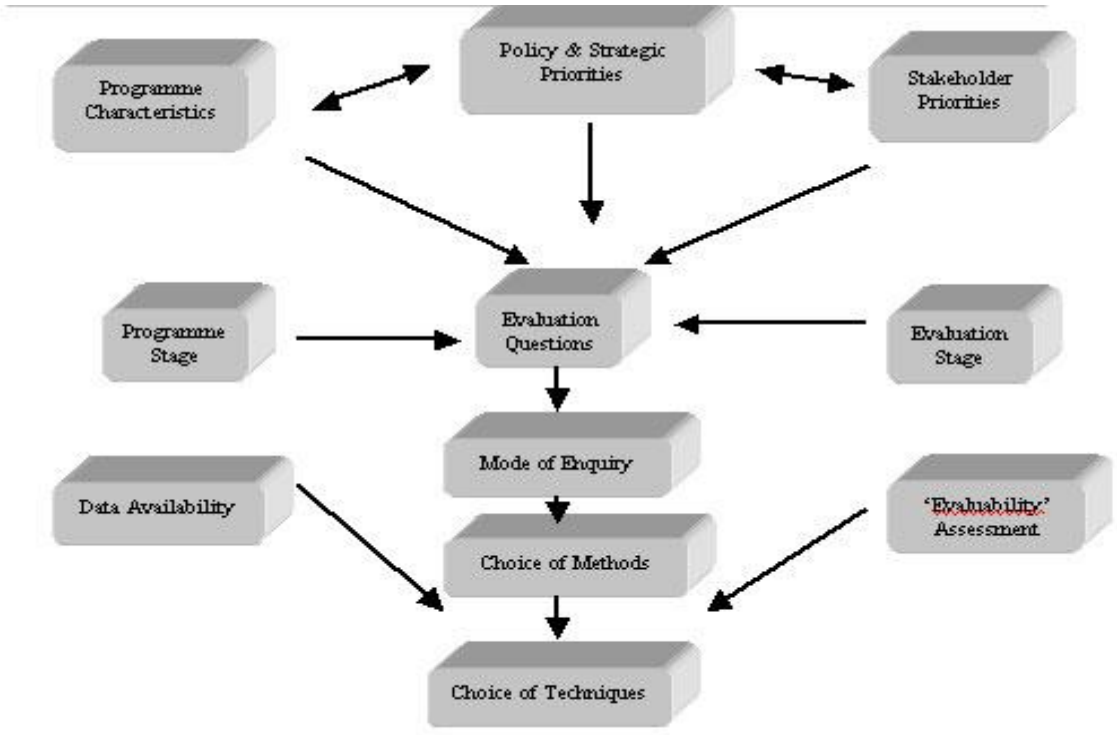
This Chapter considers the methods and techniques that are available for the evaluation of socio-economic development. The individual methods and techniques are elaborated in the Sourcebook. In Chapter 1, the design of evaluations was considered in general terms. Here we focus on the choices that need to be made both about approaches and about specific methods and techniques. The application of methods and techniques inevitably raises questions of the data sources and evidence that evaluators rely on. Here, these sources of evidence are considered not only in terms of how they can be analysed but also in terms of how they can be located or even created where they have not previously existed.

Factors influencing the choice

1) Choosing methods and techniques

As been discussed above, there are many decisions that have to be taken when designing evaluations. Stakeholders have to be consulted, intervention logics mapped out, evaluation questions identified, criteria chosen and the evaluability of what is proposed needs to be assessed. Choosing methods and techniques therefore comes some way down the line. The diagram below positions the choice of methods and techniques within this broader context.

Box Choosing Methods in a Wider Context:



Methods follow from the choice of an evaluation design or mode of enquiry: they need to answer certain kinds of questions and should only be selected if they are capable of answering these questions. This may sound obvious but one of the problems in evaluation practice is the tendency for evaluators and commissioners to favour certain methods quite apart from whether these are suitable

in terms of generating answers to the questions being asked. For various reasons, those commissioning evaluation do not always leave scope for evaluators to specify their preferred method or approach, and indeed for the early stages of an evaluation to allow for an inception report which would review and elaborate on the design, method, techniques, data collection, etc.. Nonetheless, we assume in this Chapter that evaluators will have scope to choose appropriate methods and techniques, and that the commissioners of evaluation will be informed by similar criteria and understanding as to what methods are suitable for which purpose.

Once a broad evaluation design has been chosen, the choice of specific methods and techniques still has to take account of policy and programme realities and a host of contextual factors. For example:

- The form of the socio-economic intervention: the characteristics of an evaluation object, in this case some kind of socio-economic development intervention, are important determinants of evaluation design.
- Type of evaluation purpose: Evaluations can have quite different purposes ranging from accountability through to improving management and finding out what works in what circumstances. These different purposes associated with different forms of enquiry also have implications for the methods and techniques chosen.
- Timing of the evaluation: There are different requirements for evaluation at an early, ex-ante stage, interim stage and ex-post. Each of these stages can require different methods and techniques.
- Different stages in the evaluation process. Within a single evaluation there will be the need to design, obtain data, analyse findings and draw conclusions. Each of these activities will be associated with particular methods and techniques.

This is not to suggest that there will be a one-to-one correlation between these different circumstances and contexts that evaluators will face. However, there are certain clusters of methods and techniques associated with the contexts noted above which can serve as useful points of orientation for evaluators and those who commission evaluation.

Two notes of caution are necessary:

- All methods and techniques have both strengths and weaknesses; often they are used in circumstances that are far from ideal for their application. For any evaluation, the techniques should be chosen and applied in a manner that exploits their virtues and recognises their limitations.
- Following from the above, it is best to apply methods and techniques in combination as part of any particular evaluation assignments. Relying on a single method or technique will be weaker than obtaining multiple perspectives (sometimes called triangulation).

2) Quantitative versus Qualitative : A false debate?

A common-sense distinction is often made between quantitative and qualitative methods. In fact this distinction is not as clear-cut as first appears. When qualitative statements by individuals are classified and added they become quantitative: such as, '50% of those interviewed said they had benefited from the programme'. The foundations of many quantitative evaluations are qualitative. Analyses of administrative records will often require qualitative judgements as to how to classify, for example, large, small or medium sized enterprises. Postal surveys similarly aggregate qualitative data.

However, we should not under-estimate the rigour, often embodied in widely accepted analytical protocols, that is required to convert qualitative inputs into quantitative outputs. This is why sampling, statistical significance and distinctions between different types of measurement data, among many other conventions, are critical for good quantitative evaluations.

A further blurring of the boundary between quantitative and qualitative methods follows when we distinguish between methods to gather data and methods to analyse them. Data gathered can be qualitative, e.g., interviews, surveys, focus groups and still be analysed quantitatively. Many statistical models for example use interview or questionnaire data as inputs. And quantitative analyses may only be fully understood after qualitative interpretations of what the results mean.

The nature of socio-economic development in a European context is usually bottom-up, with a mix of interventions tailored to the specific needs of territories or sectors, and is difficult to describe in standard categories. This places a limit on quantitative evaluations that attempt to provide simple comparative measures (typically indicators) or counts of results and effects. The application of indicators or other standardised measures will not be able to provide comparisons across such diverse interventions. Because of the highly contextualised nature of socio-economic development, the most effective quantitative methods will be statistical models and techniques that recognise the importance of context. For example, these need to take as their basis for comparison one setting over time (i.e., this territory in 2006 and then in 2015) and one setting in its regional context (or possibly with other matched territorial units). Such techniques - usually in the form of predictive models, macro-economic models or multivariate analyses of outcomes - should be able to assess differences between the actual and expected results of development. They should be able to answer questions such as: are trends in employment or the productivity of firms different in programme areas from other comparative areas? The use of comparative data is also important in terms of measurement of displacement: positive effects in one setting being at the expense of another. For example, has development in one territory simply sucked innovation and growth or new opportunities for marginalised groups from neighbouring territories?

In summary the strengths of quantitative evaluation are that it can:

- *Allow aggregate judgements to be made*. Policy makers want to know whether a policy or programme is working. Aggregate results, such as whether have more people got jobs across say 100 programme sites, will provide material that will support such judgements. However, aggregate measurements will not necessarily be able to demonstrate that the programmes or particular aspects of them are responsible for these changes.
- *Allow explanatory or predictive modelling*. Various sophisticated statistical and modelling techniques are useful in evaluations mainly in order to explain or predict, though less frequently to establish the causal patterns that underpin differences. So experimental methods and macro-economic models rely on quantitative data but as stressed above the methods that are generally suitable are those that take context into account.
- *Provide an overview*, which informs follow-up qualitative analysis. On the basis of global aggregate descriptive measurement it becomes clearer where sub-categories and unexpected cases occur. This directs attention towards a second, often qualitative analysis, stage.
- *Allow for estimates of extent and scale to be made*. When suitable data are available quantitative evaluation will allow calculations to be made about how much change has occurred because of an intervention. This is important especially when judgements need to be made about whether benefits are commensurate with the costs of inputs.
- *Permit some degree of comparison to be made across settings*. Policy makers need to understand whether there are different degrees of policy effectiveness across sites of intervention. Basic comparisons become easier if these can be quantified although in socio-economic development only weak forms of quantification may be possible unless supported by statistical analyses and modelling.
- *Permit stronger evaluations to be made of particular interventions*. The most effective quantitative evaluations of socio-economic development often focus on particular interventions, which are looked at separately from the wider, multi-intervention development process. So quantitative

evaluations of incentives to firms or of labour market interventions will yield strong results in terms of the outcomes of these particular interventions.

- *Allow for trend analyses to be made over time*. Quantitative measurements over time for example by gathering standard indicators on a regular basis - can help monitor changes and allow for the process of development to be tracked.

Some methods and techniques are more obviously qualitative or at least are commonly associated with the gathering of qualitative data. Interviews; participant observation or ethnographic studies; self-report diaries; discourse or content analysis of texts; rich descriptions of a local context intended to communicate a mood or ethos would all fall into that category. So also would composite methods such as case studies, which tend to draw on most of the specific techniques just referred to. In this section many of these qualitative methods are referred to and more appear in the Sourcebook. However the overriding logic behind the choice of methods is not the supposed superiority of one kind of technique or another; rather it is fitness for purpose.

Qualitative methods for gathering and analysing data are important in socio-economic development because:

- *We are interested in explaining causal patterns.* In order to learn from and replicate development, we need to understand what happens inside the black box, to go beyond inputs and outputs. Otherwise we may know what works but not how or why it works. This requires qualitative analysis.
- *We are interested in impacts for different groups.* Programmes often have different impacts for different groups of intended beneficiaries. Breaking down aggregated populations into often quite small groups allows us to investigate these differential impacts.
- *We are interested in subtle processes*. The quality of job opportunities, the experience of discrimination, a disposition towards innovation, the effectiveness of partnerships. These are subtle, qualitative phenomena that need to be captured in similarly fine-grained ways.
- *We are interested in contexts.* These are made up of many different factors, geography, history, culture, economic structures, social groups, institutional arrangements, climate, employment patterns, past development histories, etc., and the way they interact in particular development settings can only be described in qualitative terms. Furthermore the entire development process needs to be set into context if lessons are to be learned that will be transferable.
- *We are interested in human judgements*. These may be the judgements of stakeholders whose intervention logics and programme theories evaluators want to elicit. Or they may be the judgements and experiences of the intended beneficiaries of socio-economic development.
- *We are interested in bottom-up understandings*. These can include: the development ambitions of grass-roots actors (small firms, municipal authorities, professional associations) and the expectations and experiences of local people in a local development setting.
- *We are interested in innovative categories*. Development is often uncertain because it is trying to do something new. Only by examining the specific details of what is happening in a development setting will it be possible to identify the relevant categories that evaluators will need to focus on. Even if eventually the evaluation uses quantitative methods an earlier exploratory stage to clarify the relevant categories will have to come first.

3) Obtaining and using data and evidence

As was clear in Chapter 1, there can be very different views among evaluators as to the nature of evidence and what constitutes valid evidence. Those who believe in the power of objective observations (e.g., positivists) will have a different view of evidence and data than those who are more

concerned with the way perceptions and theory influence observations (e.g., constructivists). Here we take a pragmatic view. We have already acknowledged a disposition towards a realist frame of reference which, whilst valuing observation, empirical investigation and measurement when this is practical, is also concerned with the different contexts in which phenomena occur and the theories that are used to explain these phenomena. At the same time, and in certain settings, we have also acknowledged the importance of constructivist thinking, when the experience, interests and judgements of programme participants has to be given priority. Nor have we completely discarded some of the hard-won lessons of positivist science and research with its emphasis on systematic enquiry and cautious interpretation of evidence.

Scientists like to use the term data and distinctions are often made between data, i.e., the raw material of investigations, and information which has to be processed to be useful. Evidence takes us a stage further in refining information into a form that can be relied upon or is seen as strong enough to be taken seriously by users such as policy makers and programme managers.

Evaluators use very different types of data. Some data pre-exists an evaluation and will come from administrative sources (e.g., the records of a local public employment bureau or tax returns for an area). A programme through its monitoring activities will generate other data sources. (Indeed the quality of monitoring systems that are primarily the responsibility of programme managers is crucial for successful evaluation). However, some data will need to be generated by evaluators themselves, for example, by modifying monitoring systems or interviewing local SME managers or analysing the content of advertisements in trade magazines.

The quality of many evaluations would be improved if more attention was paid to using all the sources of data available. However, those who manage programmes and make policies also need to be aware of their obligation to put in place and make available sources of data that can be useful for evaluators. Nor can this be left to the point in time when an evaluation is commissioned. Putting in place basic data systems should be part of the planning of programmes as well as evaluations.

Choices for different types of interventions

Particular approaches to evaluation can tend to be associated with particular policy areas. This applies equally to the choice of methods and techniques. For example:

- Evaluation of transport interventions may include investment in infrastructure for which cost-benefit analysis and other techniques that model or describe the outcomes of different allocations of resources will be appropriate. The usage of transport systems is likely to be best covered by analysing administrative records held by transport authorities or providers. Passenger satisfaction, on the other hand, is more likely to be captured by surveys of samples of transport users.
- Evaluation of environment and energy management interventions. Here again cost benefit analyses are likely to be used at the investment stage. This is also an area where there are typically trade-offs between different priorities, for example, environmental benefits versus employment benefits. Describing and measuring such trade-offs would normally be covered by what is called environmental impact analysis. Because many aspects of environmental policy have a strong international dimension (UN and Kyoto targets) benchmarking against these international standards is also common.
- Evaluation of active labour market policies and education and training interventions make use of beneficiary surveys and panel or cohort studies that can track both the short and long term impacts of programmes. Evaluations in these policy areas also make use of experimental methods.

There is a tendency for evaluators who work intensively in particular policy areas to become wedded to particular methods and techniques. This can be for sound reasons as when this follows from the nature of the content of the policy concerned. However, this can be because these evaluators work exclusively in a particular policy area with particular evaluation and policy colleagues and tend to become isolated from the broader evaluation community and ignore ranges of methods that may be useful but with which they are not familiar.

Choices for different evaluation purposes

Another set of considerations that need to inform the selection of methods and techniques is the different evaluation purposes that were identified earlier:

- *Planning/efficiency* - ensuring that there is a justification for a policy/programme and that resources are efficiently deployed.
- *Accountability* - demonstrating how far a programme has achieved its objectives and how well it has used its resources.
- *Implementation* - improving the performance of programmes and the effectiveness of how they are delivered and managed.
- *Knowledge production* - increasing our understanding of what works in what circumstances and how different measures and interventions can be made more effective.
- *Institutional and network strengthening* - improving and developing capacity among programme participants and their networks and institutions.

To an extent, particular methods and techniques are associated with these different purposes. For example:

- With regard to *planning and efficiency*, methods are primarily concerned with resource allocation and economic efficiency. Various forms of impact analysis will be appropriate, as will different forms of cost-benefit analysis. In broader managerial terms, objective-driven techniques such as those characteristic of some logical framework approaches will also be used.
- With regard to *accountability*, methods are primarily about judging performance against some standard or target and applying relevant criteria for success and performance. In its most straightforward form, this is close to what is classically the work of auditors. Comparisons against standards can be achieved in a number of ways. For example, indicators can be used to compare actual outcomes with expectations (although here it is essential to evaluate the appropriateness of indicators as expressions of objectives and the level of the targets set). Comparisons can also be made with external examples through benchmarking. Where there is no easy way to compare externally, comparisons may be made on a before and after basis. In general the evaluations that are largely about accountability will tend to emphasise financial and monetary measures and quantitative techniques. However, this is not always so, as policy makers often find it helpful to have illustrative material and qualitative descriptions of development outcomes also.
- With regard to *implementation*, typical methods will attempt to describe processes and interim outcomes, in order to provide feedback to those responsible for programme implementation. Many of these methods and techniques will be informed by an organisational and policy studies background. There may be comparisons made between the performance of different administrative units, for example, are different regions or municipalities making more or less progress? Case studies of organisational and partnership arrangements will help understand the strengths and weaknesses of different implementation approaches. Often these kinds of methods will involve what are called formative evaluation methods and techniques. These

place a particular onus on the evaluator to provide feedback in ways that will be useful and will help programme managers translate emerging evidence into practical action.

- With regard to *knowledge production*, methods will be closest to those used by academic researchers. They will be subject to demands for rigour, representativeness and the cautious interpretation of findings, especially where these may be inconsistent. Typically, for knowledge production purposes, evaluators will want to answer the question, what works? This would be an area where experimental methods are seen as relevant. However, the diverse and bottom-up nature of socio-economic interventions, the way these are combined in particular configurations and the different localities and contexts where programmes take place, makes traditional experiments challenging to apply. For that reason that realist thinking, with its emphasis on the influence of context on outcomes, should be integrated in the research. Here the more complex question is asked: what works, for whom, how and in what circumstances? Methods and techniques suitable for this will generally involve comparison between different cases selected to demonstrate alternative interventions and alternative contexts. Such comparisons may be based on case studies, data-bases that structure intervention/outcome/context configurations or a range of other techniques that are able to capture and describe these different aspects of socio-economic development.

It is widely accepted in the evaluation community that reliable knowledge rarely comes from a single evaluation. For this reason there is growing interest in undertaking synthesis studies and various kinds of meta-analysis that try to build up what is known from a range evaluations. Such analysis is strengthened if, when designing evaluations that might subsequently be the included in meta-analyses, some standard structures and data items are collected across all cases.

With regard to *institutional and network strengthening*, it is now widely recognised that evaluations are not exclusively to meet the needs of programme managers and sponsors but also have to be owned by a wide group of stakeholders. Furthermore, the effective delivery of programmes often depends on the capacities of the institutions and organisations from which these stakeholders come, as well as broader civil society networks. Very often the methods that would be appropriate in these settings will be participatory: placing an emphasis on close collaborative work between the evaluators and the institutions and networks involved. These participatory approaches will not only be important in formulating evaluation questions but also when generating data and using these results of evaluations. For example, in a community setting where there are many interests and perhaps a lack of a shared view, evaluators may need to work with community representatives to develop consensus if the results of an evaluation are to be used.

Choices for different programme/policy stages

The importance of the time-cycle in programmes and policies has been a theme throughout this GUIDE. In European Cohesion Policy this is formalised in terms of ex-ante, interim and ex-post evaluations. Quite apart from these particular labels, the underlying stages of a programme from policy formulation and programme design through to implementation and delivery and conclusion or results and impact poses certain demands for evaluation in most major programmes:

At the *design* stage, there will be an emphasis on identifying appropriate interventions and the organisation management arrangements able to deliver them;

At the *implementation* stage, there will be an emphasis on feedback processes, first outputs and results and providing feedback in a way that supports learning;

At the *conclusions or results* stage, there will be an emphasis on results and impacts for intended beneficiaries or territories in relation to intentions (e.g., following from objectives) as well as

unintended consequences. For 2014-2020, the conclusions stage starts well before the programmes are completed, as programme authorities must provide evaluations of how the various interventions are contributing to their objectives during the programming period at a time when they decide is most appropriate. For the European Commission, the ex post evaluation is often required to be completed at the same point in time as spending stops.

1) Design: Interventions and organisation

Policy and programme formulation usually entails the identification of starting circumstances and of desired objectives. However, the links in the chain that connect present circumstances with a desired future will not be specified. This is what happens at the design stage. Constructing programme theories or logic models of socio-economic programmes showing the implementation chains associated with particular interventions is a useful way of filling out the stages between baseline circumstances and longer-term intended change. The use of these kinds of models can be supplemented by other techniques such as evaluability assessment, which go beyond logical frameworks to actively involve programme managers and policy makers in assessing what can be delivered in a feasible way.

Again we need to be aware of the role of many stakeholders in socio-economic development programmes. This makes it useful to combine programme design with participative methods that can also begin to shape later stages in an evaluation. Actively involving groups of stakeholders in putting together their own programme theory rather than relying on a single exercise with policy makers can be one approach. Some forms of programme theory building such as the so-called theory of change approach are designed to be participatory and to elicit the understandings and implicit theories of stakeholders as actors rather than as mere recipients of programme inputs.

Evaluation techniques need to be applied at the programmatic level as well as in terms of individual interventions. For example, project appraisal techniques including cost-benefit analysis can be used to inform choices between different interventions intended to achieve the same objectives. It may also be useful to assess the trade-offs between different measures and interventions.

A crucially important contribution of evaluation at this stage is to make explicit the logic of intervention and assess the plausibility of the amount of resources delivering certain outputs under certain policy actions making a contribution to results.

Synthesis studies of previous implementation mechanisms can also be undertaken at this stage. For example, what is known about suitable organisational and administrative arrangements? What kinds of decision making and partnership architecture will be most effective? These kinds of questions are probably best answered by comparative case studies and literature reviews of existing evaluations.

2) Implementation: Feedback and intermediate outcomes

Throughout the implementation of a programme there is a need for feedback to allow programme managers to identify problems and take remedial action. Monitoring systems will provide much of this feedback. However, monitoring systems themselves may help identify problem areas that deserve more detailed investigation. For example, slow start-up in particular projects and consequent under-spending of budgets or the withdrawal of support from an important group of stakeholders may justify these kinds of evaluation activities.

When the intervention being planned is particularly innovative or experimental, there may be a justification for tracking or following the implementation process in some detail. Such formative evaluation activities are likely to involve techniques such as participant observation which would need to be reinforced by systematic feedback. At this stage in the process feedback can be very welcome but can also be quite threatening. Various kinds of communication and consultation skills

are needed in order to manage this kind of feedback in a constructive way. This kind of evaluation can also demand skills from programme managers; for example, they may need to conduct their own self-evaluations as part of an overall evaluation process.

Monitoring systems will also track intermediate outcomes. Assuming that logical frameworks and programme theories have been constructed thoroughly at the design stage, a template should exist that will describe the milestones expected at different programme stages.

3) Conclusions: Results and impacts

Policy makers for accountability reasons and key stakeholders because of their own needs and commitments look to evaluation to provide information on results and estimates of impacts at the end of a programme cycle. Evaluation methods will seek to compare what has been achieved with what was intended and endpoints with baselines, including an assessment of the contribution of the policy to the results achieved. A broad range of techniques can be deployed including:

- Surveys of intended beneficiaries,
- Econometric or statistical models to demonstrate changes in economic performance compared with predicted results (perhaps by comparing trends in a development setting with other settings and using models developed at the beginning of a development cycle),
- Literature review of similar policies,
- Control and comparison groups,
- Participatory methods including workshops and focus groups,
- Case studies, and
- Indicators based on contextual data or administrative data provided by public authorities.

Acquiring and using data in evaluation

1) All data are "produced"

Evaluators depend on data: the raw material that once collected is organised, described, grouped, counted and manipulated by various methods and techniques. Distinctions are often drawn between data that are primary - generated as a direct consequence of a programme or intervention - and secondary - generated for other purposes and pre-exist the programme or intervention. For example secondary data sources might include:

- Statistical sources such as national and regional statistics and EUROSTAT[3],
- Annual reports of development authorities or federations of enterprises, and
- Administrative records of public employment agencies, taxation returns, qualifications and training data.

None of this data happens without considerable effort and evaluators need to know how secondary data was put together before using them. What samples were used, how were outcomes defined, what is the timescale covered, what is the unit of analysis? It is only by asking these questions that a judgement can be made about their usability in a particular evaluation.

It is easier for an evaluator to understand the provenance of primary data. These can include:

[3] http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/

- Monitoring data produced by a programme as part of its reporting obligations to funding authorities,
- Usage data generated by the use or uptake of services, funds or facilities provided by a programme, and
- Data collected from development sites and intended beneficiaries by evaluators through surveys of beneficiaries, counts of those using a consultancy fund, focus groups and stakeholder consultations.

However, here also data do not emerge fully formed. Their collection has also involved the application of protocols and techniques that specify what they can be used for. Does usage data differentiate between different types of users? Is monitoring information confined to financial data? How representative are samples of beneficiaries?

Because all data are processed and is the result of decisions made in the course of collection, evaluators need to understand what these decisions were, especially when these decisions were made by others, as with secondary data. This is not always easy. With regard to primary data which are generated by or close to a programme and its evaluation - the evaluation team is better placed to know what decisions were made. Even here there will be a distinction between those forms of data that are directly produced by an evaluation and those that are generated by the programme. However even when data are collected directly by the evaluation team its strengths, limits, scope and relevance need to be thought through in terms of the kinds of future analyses that will be made and the kinds of arguments that the data will be expected to support.

This is a further argument for thinking through evaluation at the design stage with care. The collection of data needs to be considered in tandem with the choice of methods for analysis.

2) Accessing data as a planned activity

Multiple interventions involving many partners mean that data that concern any single programme will be held in many different places. Simply mapping out where data is held and what is available is a serious task. Negotiating and agreeing the terms under which data will be provided or made available can be more complex. Administrative data in particular can be the subject of various confidentiality or data protection rules. Sometimes administrative data can only be released when identification labels (names and postcodes) are eliminated. Even when this is not a problem, administrative bodies are often jealous about their information sources. Negotiating access to data is a task to which time always needs to be allocated.

Sometimes the process of accessing data sources can itself generate useful data for an evaluation. For example the willingness of partners to share information can be regarded as an indicator of the coherence and strength of a partnership. Constant refusal to share information, for example, suggests that the partnership is not characterised by high levels of trust.

When evaluators are directly involved in generating data as with many primary data sources problems of access still exist and need to be considered carefully. Examples of access issues can include the following.

Ensuring high levels of response in sample surveys. Low response rates are far too frequent in evaluations and this can weaken the evidence base for conclusions and judgements. There are many possible ways of improving response rates:

- Communicating (perhaps in the local or trade press) clearly what is the purpose of an evaluation in general and of surveys in particular;
- Designing survey instruments in simple non technical language and keeping the surveys short;

- Devoting time to follow-up activities reminder letters, phone calls and re-issue of survey instruments after an elapse time to known non-respondents.

Getting access to disadvantaged groups. What are sometimes called hard to reach groups are often critical for an evaluation. Such groups may be distrustful of official action, and this may carry over to an evaluation. Ways of overcoming these problems can include:

- Making links with community gatekeepers so that they can act as local advocates of the evaluation.
- Producing instruments in different languages (when minority languages are used) or in multiple formats - Braille or audio tapes for those with disabilities.
- Employing local people from these groups to collect information, run focus groups and explain the evaluation within their own networks.

Beneficiaries and stakeholders wanting to get something out of an evaluation. A major reason for non-cooperation - or less than enthusiastic cooperation - is a sense that those being asked to cooperate will get no benefits from the exercise. This can be overcome or at least reduced if:

- There is adequate involvement of beneficiaries and stakeholders in the design stages of the overall evaluation and in designing and piloting particular instruments.
- Guarantees are given that all those cooperating will receive feedback. This can take the form of a publicly available report, a feedback letter containing an executive summary or an invitation to a feedback meeting once the evaluation is complete.

The quality of data and the willingness of gatekeepers, stakeholders and beneficiaries to cooperate will be a key determinant of data quality and ultimately of the evaluation. It is worth devoting attention to data quality and planning as an integral part of gathering data and choosing methods.

3) Quantitative and Qualitative data

We have emphasised the importance of drawing on a full range of evaluation methods and techniques, including those that are both quantitative and qualitative. Here we briefly highlight some of the characteristics of data, in the terms of their qualitative and quantitative distinctions.

First the importance of distinguishing between data as collected and data as analysed is important to reiterate. As has already been noted virtually all data need to be produced and processed to become usable. However even when various methods of analysis have been applied, there will be differences in the characteristics and strengths of quantitative data.

What is called quantitative data can take very different forms, for example:

- It may be nothing more than a way of describing categories. Categoric or nominal data has no numeric value, rather numbers are used to distinguish different categories. Thus Group 1, 2, 3 and 4 may be labels applied to four different sets of SMEs distinguished by the sectors in which they operate.
- Slightly stronger in terms of quantification can be ordinal data where it is known that some items are more or less, bigger or smaller, than each other. For example, some firms may be growing and some declining and this can be used as a source of data even if one does not calibrate or have access to the precise differences between the growth and decline of each firm.
- A still stronger form of quantification would occur when one can place relative differences on a scale where the intervals between the scale can be known. Various scoring and ranking systems and items in questionnaires would conform to this kind of quantification. For example, an expert might rate the environmental impact of a project as anything from -3 to +3

or a questionnaire respondent might be invited to position herself on a scale of satisfaction from 1-5. Even though these forms of quantification are stronger than those described previously, they are relatively weak in terms of their numerical and calculative possibilities.

- Ratio data is the strongest form of quantification. This occurs when there is a known zero point on a scale. So one is not dealing with an invented series of intervals but rather with something that can be measured independently. For example, monetary values, age profiles of populations, export flows and productivity indices based on annual output records would usually be a form of ratio data. Arguably this is what is usually meant when we speak of quantitative data even though it is less common in evaluation than we sometimes imagine.

The justification for this relatively technical description of different strengths of quantitative data is to emphasise that in socio-economic development, most so-called quantitative data is, in fact, relatively weak. Reductions in social exclusion, improvements in human resource quality and diversification of a rural economy will usually depend on categoric, ordinal and interval data. Such measures cannot be considered objective or as strong as it is sometimes assumed quantitative data always are. There will be some kinds of data, for example, data related to the competitiveness of local firms or participation in the labour market, which can be subject to more rigorous analysis using what is described above as ratio data. However, such levels of rigour will be less frequent than sometimes assumed.

As the above distinctions suggest quantitative/qualitative data can be best be understood as a continuum from the most quantitative to the most qualitative. In many ways categoric and ordinal data can be seen as relatively qualitative. What might be regarded as pure qualitative data is highly diverse, perhaps made up of unique instances or reports and only able to be described on a case-by-case basis. A case study or a life history that was put together without any comparison or prior categorisation might conform to such a qualitative ideal. However, as soon as several instances of such cases or biographies are collected, the same process of categorisation becomes possible as has been described above under the quantitative heading.

In qualitative terms data can be compared and thus eventually analysed along a number of dimensions. Categorisations or ranking along a scale is most appropriate where one is dealing with a population of related phenomena, for example, a number of individuals, a sample of firms, or a sample of households. However, when one is dealing with unique or individual examples such as a particular rural community or a particular manufacturing sector, comparisons are more likely to occur over time (before and after measures) or in relation to some external standard or criterion.

The end-point of this kind of perspective is to blur the quantitative/qualitative distinction. The distinction is stronger in terms of analytic intent. Peoples' views and opinions can be unique and qualitative or typified within a broader classification: but the raw data remains the same. It is a question of how raw data is processed for the purposes of analysis. Quantitative data are most likely to be used when aggregation and generalisation is required; and qualitative data when complexity and needs to be described. The choice between such strategies must ultimately depend on what questions an evaluation is trying to answer. Ultimately, most evaluations will need both quantitative and qualitative data.

Creating indicators and indicator systems

1) What are indicators?

Definition and characteristics of an indicator

An indicator can be defined as the measurement of an objective to be met, a resource mobilised, an effect obtained, a gauge of quality or a context variable. An indicator produces quantified information with a view to helping actors concerned with public interventions to communicate, negotiate or make

decisions. Within the framework of evaluation, the most important indicators are linked to the success criteria of public interventions.

In order to be useful it is preferable if an indicator has the following characteristics:

- The indicator definition is closely linked to a policy objective. (Indicators are most helpful when objectives have been specified in terms of targets or milestones that apply to the definition of the indicator).
- The indicator is measured regularly. It is helpful to have time series information where the precise indicator definitions have been applied consistently. Ideally data should be available from prior to the adoption or implementation of the intervention. However, interventions often themselves call for new data to be collected.
- Steps are taken to ensure data gathered is reliable. For example, for output indicators where data are provided directly by the projects, sample checks should verify the data. For result indicators, it is probably better that data are gathered on an independent basis.

In practice indicators rarely exhibit all of these characteristics and it is likely to be necessary to gather evidence from a variety of disparate sources including:

- The inputs to and timing of the programming process;
- Secondary sources;
- Primary sources, including Stakeholder surveys;
- Administrative information
- Statistics.

Much of this information may have been gathered for purposes other than evaluation.

An indicator quantifies an element considered to be relevant to the monitoring or evaluation of a programme. For example: "1200 long-term unemployed received training financed by the programme" or "75% of the participants in training courses claim to be satisfied or highly satisfied".

A good indicator should provide simple information that both the supplier and the user can easily communicate and understand. The following are examples of indicators that are readily understood: rate of budget absorption; percentage of regional firms assisted; number of gross jobs supported; and number of jobless in the eligible area.

An indicator may have several values over time. The unemployment rate, for example, may have a different value at the outset from a value taken mid-way through the implementation of a programme, and so on. Variations over time constitute trends.

Type of indicators

There are several typologies of indicators:

- In relation to variables: Complete, partial and complex indicators
- In relation to the processing of information: Elementary, derived and compound indicators
- In relation to the comparability of information: Specific and common indicators
- In relation to the scope of information: Context and programme indicators
- In relation to the phases of completion of the programme: Resource and output indicators
- In relation to evaluation criteria: Relevance, efficiency, and effectiveness indicators

The most useful of these typologies for socio-economic programmes is the distinction between resources (inputs); outputs and results indicators.

Resource indicators provide information on the financial, human, material, organisational or regulatory means used by to implement programmes. Resources are the joint responsibility of the financing authorities, which allocate them, and the operators who implement them. Most resource indicators are regularly quantified by monitoring systems. Examples of resource indicators include: the total budget (quantity of resources); annual budget absorption (resource absorption rate); percentage of expected over/under spending; percentage of European financing in the total public financing; number of people working on the implementation of the programme; number of organisations involved in the implementation.

Output indicators represent the product of the programmes activity. More precisely, an output is considered to be everything that is obtained in exchange for public expenditure. Outputs are normally under the entire responsibility of operators who report on them through the monitoring system. Examples of output indicators include: kilometres of roads built; hectares of urban wasteland rehabilitated; capacity of purification plants built; number of trainees whose training was paid by the programme.

Result indicators represent change sought by the programme or priority to which the programme or priority should contribute. Readers' attention is drawn to the changing concepts of "results" and "impacts" in the evaluation literature[4]. For further reading material read the guidance produced by the DGs for Regional Policy and Employment and Social Affairs.

Impact indicators have been removed from Structural Funds programmes, except those dealing with fisheries and rural development. For the DGs for Regional Policy and Employment and Social Affairs impact is the contribution of the intervention to change. It cannot be monitored; it requires evaluation.

2) Indicators and evaluation

Indicators serve a number of useful roles in evaluation. The use of indicators normally forms part of an evaluation particularly where objectives are expressed in clear operational terms. The information they provide needs to be carefully interpreted in the light of other evidence in order that evaluative conclusions can be drawn. Indicators have the potential to contribute to the evaluation of socio economic programmes in several ways:

- The analysis of the indicator scores can be used to provide support for a rationale for intervention and resource allocation.
- Indicators can be used to compare inputs and outputs in order to measure efficiency.
- Indicators can be used to compare actual outcomes with expectations in order to assess how needs have been addressed.
- Indicators can be used to compare inputs relative to impacts and hence allow the assessment of the value (value added) of policy, legislation or initiatives.

The system of indicators and the programme cycle

Indicators are used at the beginning of the programme cycle to help to define territories eligible for assistance, to analyse the regional context, to diagnose economic and social problems to be addressed, and to assess the needs that the programme has to meet. At this stage, indicators such as the number of start-ups, unemployment rate or disparities between infrastructures often play a decisive role.

[4] See: http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf
And http://evi.sagepub.com/content/16/2/153.abstract

The choice and validation of the intervention strategy constitute the second stage in the programming cycle. At this stage the programme designers should define the objectives precisely and quantify them. The specific objective of a priority should have a result indicator defining the change sought. Baselines should be known. The intervention logic should plausibly demonstrate that the amount of resources and outputs can deliver results.

Once defined and adopted, the programme is implemented. It is monitored and evaluated on an on going basis. Indicators serve to monitor the pace at which budgets are spent, the extent to which the schedule is adhered to, the proportion of the eligible population reached, the progress in achieving the planned outputs of the programme. During the evaluation the programme managers carry out evaluation to assess the extent to which the interventions are contributing to their specific objectives and the result indicator.

The programming cycle ends with an ex post evaluation, one of the main functions of which is to report on programme or policy results and on the extent to which aims have been achieved. The use of indicators is essential at this stage to communicate the achievements of the programme.

Organisational aspects: Involving users and suppliers of information

A system of indicators has more chance of functioning when the suppliers and users of the information have been involved in its creation. In contrast, a closed group of specialists will be tempted to construct an expensive, technically ideal system that may never operate satisfactorily.

As far as the users are concerned, explicit support from the highest level of the authority managing the programme has to be assured. It is then advisable to create a group of future users of the system, and to give it the job of defining the indicators.

The main suppliers of information are the operators who implement the projects in the field. Their participation is likely to ensure that the system is pragmatic because they are familiar with the practical possibilities and limits of data collection.

3) Selecting indicators

Selection of the most relevant indicators

Each of the programme actors have their own responsibilities, their own areas of decision-making and their own information needs. As a result, all indicators are not useful at all levels. On the contrary, it is generally accepted that each actor requires an operating report with a small number of indicators, selected as the most relevant in relation to the nature of the decisions that have to be made. It has been shown that in a situation of decision-making, a person cannot take into account more than about ten indicators at once. When there are too many indicators decision-makers are swamped with an excess of information.

The heterogeneity of programmes

The experience of Cohesion Policy has shown that it is difficult to choose indicators that are absolutely necessary for the monitoring and evaluation of a programme. Because the programmes are multi-sectoral and multi-objective, there is a tendency to want to measure everything and to design systems of indicators that are so heavy that it is impossible to make them work. In practice, it is impossible to produce and regularly use such a large amount of information.

For the 2014-2020 period, there is a major emphasis on reducing the number of indicators, which is also linked to a need to concentrate resources. Indicators should feed policy debate; too many do not communicate the achievements or not of policy.

4) Avoiding the adverse effects of indicators

The use of indicators is often hindered by the fear of provoking adverse effects. There are several types of adverse affect:

- The skimming-off or creaming effect,
- Convergence to the average,
- Unanticipated effects where results are subordinated to indicator scores.

Skimming-off effects can occur when the performance of training and employment services organisations is measured by the placement rate of beneficiaries. To obtain a better placement rate for their beneficiaries, it is in the organisations' interests to recruit people in the best possible situation who also meet the eligibility criteria. The operators therefore tend to "skim off" or favour those beneficiaries whose employability is higher. This effect is undesirable because it focuses assistance on those who are in relatively less need.

An example of how indicators caused a reduction in differences by a convergence towards the average is given below. An indicator can also encourage behaviour leading to sub-standard performance. This occurs when the indicator rewards undesired results or when the system causes the operators to work for the indicator rather than for the result.

Convergence towards the average rather than excellence:

The British Audit Commission's system of indicators organises the quantification of about two hundred output and result indicators relating to the services offered by municipalities. The indicators are comparable from one town to the next and are published annually in the local press. This system creates a very strong impression of the town when one of the services performs badly. As a result, many municipalities increased the budgets of their services with the lowest performance, in the hope of raising the standard. In these cases, financial resources were sometimes drawn from the most effective services. Use of indicators thus caused a reduction in differences by convergence towards the average. It was an adverse effect because it was hoped that the indicators would cause performance to converge towards excellence.

Adverse effects inevitably appear after two or three years of functioning of a system of indicators, no matter how well it is designed. These undesirable effects are generally not foreseeable.

The possible appearance of adverse effects should not be an argument for refusing to measure performance. It is possible to minimise adverse effects, either by amending the indicator causing the problem, or by creating a procedure for interpretation of the indicator by expert panels. It is then important to watch out for the appearance of adverse effects and to correct the system when these effects appear.

A further final point to emphasise is that indicators do not tell you everything about performance. Evaluators, above all, must remember this. Their task is to understand and communicate why and how phenomena occur – indicators cannot tell you this. This is the core task of evaluation.

Golden rules

1. Choosing methods and techniques follows directly from the kind of questions one wants to ask and these questions are part of an extensive design exercise that includes consulting stakeholders and assessing programmes characteristics. Choosing methods and techniques first and trying to make them fit with questions for which they have not been specifically chosen will create problems. The techniques chosen need to reflect the purpose and focus of the evaluation.

2. Most techniques have strengths and weaknesses; these need to be recognised and where possible different techniques need to be applied together to strengthen the analysis and make the evaluation results and conclusions more reliable.
3. Because of the distinctive character of socio-economic development: bottom-up, using different combinations of interventions and tailored to territorial and sectoral needs, it is difficult to measure and compare outcomes across programme settings. This doesn't mean that measurement, quantification and statistics are not relevant. They can be powerful tools when comparisons are at the level of the particular development programme and do not attempt to compare non comparable settings.
4. Qualitative methods and techniques are well suited to socio-economic development because of the subtlety and holistic nature of what is being attempted and because of the differences in contexts which need to be described in qualitative ways.
5. Thematic priorities which are very common in European programmes pose real difficulties for evaluators. Because policy makers want to understand how far their policies are successful as a whole, there is often pressure to aggregate results and find a common way of describing or even measuring what is happening. This often cannot be done. Sometimes only qualitative descriptions will work. Take care not to add up apple and pears.
6. There is often a tension between choosing evaluators who know a lot about a particular policy area and those whose evaluation skills are more generic. Ideally in an evaluation team one tries to span both of these sets of knowledge and experience. Commissioners need to be aware of the dangers of contracting evaluators who have lived most of their professional lives working in one specialised area and using a limited set of methods and techniques. This is another argument for looking at the balance of skills in the evaluation team.
7. It is important to distinguish between methods and techniques for gathering data, for analysing data and for informing evaluative judgements. This distinction is not always made partly because those who undertake evaluations may be more preoccupied with one stage of the process rather than another. As in all things there needs to be a balance.
8. Data are never pure or naturally occurring, they need to be produced. Because of this evaluators need to know from where their data comes and what decisions have made in the course of their production. The strength of the arguments and conclusions that can be drawn depend on the strengths and characteristics of the data being used.
9. One important aspect in evaluating data follows from the way they have been accessed and how access has been negotiated. Different partners have to be willing to share information and all stakeholders need to be convinced that they are going to get something out of an evaluation before they give access with any enthusiasm to any information they hold. Investing in these kinds negotiation processes will make a difference to quality and the evaluation as a whole.
10. The quantitative/qualitative divide is overstated. What is needed is quantitative data to provide overviews, for example to aggregate outputs of an intervention and provide a comparative perspective, and qualitative data able to capture subtleties, people's experience and judgements.
11. Well-conceived indicators and monitoring systems are a powerful adjunct to evaluation. Often evaluators depend on monitoring systems which are indicator based. If these are not put in place in the design of the programme it will be too late to create such systems later on.
12. Over-elaborate indicator systems will be counterproductive. While there is a temptation to measure everything, this should be resisted. This can be costly and the results useless.
13. Indicators do not tell you everything about performance. Often they are a prompt to understand why and how something is happening. This is the role of evaluation.

Bibliography

Barca F, An Agenda for a Reformed Cohesion Policy, European Commission, 2009

Boyd, R., Grasper, P., Trout, J.D. The Philosophy of Science, Cambridge, MA: MIT Press, 1991

Boyle, R., Lemaire, D., eds. Building Effective Evaluation Capacity: Lessons from Practice. New Brunswick and London: Transaction Publishers, 1999

Chen, H. T: Theory- Driven Evaluations, Newbury Park, CA: Sage: 1990

Connell J. P., Kubisch, A. C., Schorr, L. B. and Weiss, C. H. eds. New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts, New York: The Aspen Institute, 1995

Fetterman, D. M., Kaftarian, S. J., Wandersman, A., Eds. Empowerment Evaluation: Knowledge and Tool for Self Assessment and Accountability, Thousand Oaks, London, New Delhi: Sage, 1996

Furubo, J.E., Rist, R., Sandhal, R. (eds), International Atlas of Evaluation, New Brunswick London, Transaction Publishers, 2002.

Guba, E. G. and Lincoln, Y. S. Fourth Generation Evaluation. Newbury Park, London, New Delhi: Sage 1989

House E. R., Assumptions Underlying Evaluation Models. In Madaus G. F., 1989

Jacob, S., Varone, F. Evaluer l'action publique: tat des lieux et perspective en Belgique, Gent, Accademia Press, 2003.

Leeuw, F. L., Rist, R. C. Can Governments Learn? Transaction Publishers, 1994

Nutley, S. M., Walter, I., Davies, H. T. O. From Knowledge to doing: A Framework for Understanding the Evidence-into-Practice Agenda. Evaluation: the International Journal of Theory, Research and Practice, 2002, Vol. 9, No. 2

Patton, M. Q. Qualitative Research and Evaluation Methods. Thousand Oaks, London, New Delhi: Sage, 2002

Pawson, R. Evidence Based Policy: In Search of a Method, Evaluation: the International Journal of Theory, Research and Practice, 2002, Vol. 8, No. 2, pp. 157-181 (a)

Pawson, R. and Tilley. N. Realistic Evaluation. London, Thousand Oaks, New Delhi: Sage, 1997

Pollitt, C., Justification by Works or by Faith? Evaluating the New Public Management: Evaluation, Vol. 1 (2): 133-154 (1995)

Pressman, J. and Wildavsky, A: Implementation. (Berkeley, University of California Press: 1973

Rist, R. C. Furubo J. E., Sandahl R. eds. The Evaluation Atlas. New Brunwick London: 2001

Rogers, E., Diffusion of Innovation. New York: The Free Press, 1995

Shadish, W., Cook, T. and Leviton, L. (eds): Foundations of Program Evaluation: Theories and Practice. Newbury Park, Sage, 1991

Van der Knaap, P. Policy Evaluation and Learning. Feedback, Enlightenment or Augmentation? Evaluation Vol. 1 (2): 189-216, 1995

White, H, A Contribuion to Current Debates in Impact Evaluation, in Evaluation, Sage, 2010

Wholey, J. Using evaluation to improve government performance' Evaluation Practice, Vol. 7, pp.5-13, 1986

Glossary A – Z

Absorption Rate

Budgetary inputs mobilised in proportion to the inputs initially allocated.

Accountability

Obligation, for the actors participating in the introduction or implementation of a public intervention, to provide political authorities and the general public with information and explanations on the expected and actual results of an intervention, with regard to the sound use of public resources.

Action research

Practice-based research, which seeks to end the dislocation of research from practice and enhance the position of research as a direct mechanism for change and improvement. Action research projects usually have the following characteristics: investigation of practices in terms of strategic actions with potential for improvement; collaborative working between evaluators and stakeholders; a methodology involving a series of interrelated cycles of planning, action, observation and reflection; those responsible for the practice are at the heart of these cycles and are the key participants in each stage.

Additionality

Literally, additionality is the extent to which something happens as a result of an intervention that would not have occurred in the absence of the intervention. In the context of EU Structural Fund, additionality means that Community support for economic and social development does not displace efforts by national governments. In other words, the fact that the beneficiary State's own financing remains, globally, at least equal to that which existed before the Structural Funds' contribution. Verification of the implementation of this principle is carried out at the national level in the context of financial control and not of evaluation as such.

Administrative data

Information relating to the administration of the Programme usually collected through a structured monitoring process. Not necessarily for the purposes of evaluation.

Aim

Broad intention or overarching objective, usually to bring about positive change.

Audit

Verification of the legality and regularity of the application of resources. Auditing makes a judgement in terms of criteria and standards, known and clarified beforehand. The main purpose of an audit is to ensure compliance. With time, the terms "control" and "audit" have been extended to encompass more activities. For example, certain audits or controls check whether the outputs have been produced with an adequate degree of efficiency and quality and whether they offer value for money (performance audit).

Average

In mathematics, an average, or central tendency of a data set refers to a measure of the "middle" or "expected" value of the data set. There are many different descriptive statistics that can be chosen as a measurement of the central tendency of the data items. An average is a single value that is meant to typify a list of values. The most common method is the arithmetic mean but there are many other

types of averages, such as median (which is used most often when the distribution of the values is skewed with some small numbers of very high values, as seen with house prices or incomes).

Baseline

Facts about the condition or performance of subjects prior to treatment. The essential result of the pre-test part of the pre-test/post-test approach. Gathering baseline data is one of the key reasons for starting an evaluation before a program starts, something that always seems odd to budgetary bureaucrats. (Scriven M., Evaluation Thesaurus)

Benchmarking

Qualitative and quantitative standard for comparison of the performance of an intervention. Such a standard will often be the best in the same domain of intervention or in a related domain. Benchmarking is facilitated when, at the national or regional level, there is comparative information of good and not so good practice. The term benchmarking is also used to refer to the comparison of contextual conditions between territories.

Beneficiary

Person or organisation directly affected by the intervention whether intended or unintended. Some people may be beneficiaries without necessarily belonging to the group targeted by the intervention. Similarly, the entire eligible group does not necessarily consist of beneficiaries.

Call for Proposals

A call for proposals is a process through which organisations are requested to submit proposals to undertake work / research that matches the requirements of a brief / research question. Organisations then submit tenders to carry out the research / project, which usually involves providing details of the proposed research methodology.

Case Study

A technique involving the examination of a limited number of specific cases or projects which the evaluator anticipates will be revealing about the programme as a whole. Case studies tend to be appropriate where it is extremely difficult to choose a sample large enough to be statistically generalizable to the population as a whole; where generalization is not important; where in-depth, usually descriptive data is required; and where the cases or projects to be studied are likely to be quite complex.

Causal analysis

The analysis that attempts to establish a relationship of cause and effect between observed phenomena. In the case of evaluation, causal analysis attempts to establish a relationship of cause and effect between a public intervention and the changes (or lack there of) observed in one or more outcomes of interest.

Causal effect

The consequence attributable to a policy or intervention. Causal analysis aims to identify the effect of the intervention on the phenomenon under study, not at unravelling all causes of the phenomenon. We employ the term "effect" only referred to the "cause" of interest--that is, the public intervention being evaluated.

Cluster analysis

A data analysis tool for solving classification problems. Its aim is to sort cases (people, things, events, etc) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters.

Colour Vote

Technique used to run meetings, based on a visual presentation of opinions in a group. In an evaluation situation in which several stakeholders are involved, it can be used to structure discussion and collective reflection by facilitating the expression of all points of view. It is therefore a tool likely to enhance the impartiality of the evaluation in the context of a pluralistic approach. It may be used to clarify and grade evaluative questions, to choose evaluation criteria, to validate conclusions, and to formulate recommendations.

Commissioner

Person or organisation which decides to launch an evaluation. Commissioners have the advantage of making their priorities explicit by establishing a mandate. A commissioner will steer the work of an evaluation team either directly or through a multi-stakeholder steering group or other evaluation authority set up for this purpose.

Comparability

Quality of an indicator which uses the same measurement unit to quantify the needs, objectives or effects of several different interventions. Comparability is useful for establishing norms for judgement (e.g. the average cost of jobs created by the intervention can be favourably compared to that of similar interventions). Efforts made to improve comparability involve the harmonisation of measurement units and result, initially, in the definition of standard indicators, i.e. indicators that can be used in several regions with the same definition for the same sector of intervention (e.g. number of SMEs assisted, defined and calculated in a comparable way).

Competitiveness

The capacity of a firm, in a competitive socio-economic environment, to defend or increase its market share over the medium and longer term, and to generate wealth. The notion of competitiveness can apply to a single firm or to all the businesses in a sector or region. It is directly related to the notion of competitive advantage, an advantage which an enterprise, region or sector must possess or develop if it is to be competitive in a segment of a particular market.

Complementarity

The fact that several public interventions (or several components of an intervention) contribute towards the achievement of the same objective.

Completion rate

Percentage of initially planned outputs which have been completed. The completion rate of a major infrastructure project is calculated in terms of the stages of work which were initially planned and were actually completed.

Concept mapping of impacts

Tool used for the clarification of underlying concepts which may include explicit and implicit objectives. It relies on the identification, grouping together and rating of expected outcomes and impacts. The concept mapping of impacts is implemented in a participatory way, so that a large

number of participants or stakeholders can be involved. It may result in the selection of indicators that are associated with the main expected impacts.

Confounding factor

A confounding factor (also confounding variable, lurking variable, a confound, or confounder) is an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable. In the presence of confounders, the relationship between two observed variables is termed a spurious relationship. Thus, confounding is a major threat to the validity of inferences made about cause and effect, i.e. internal validity.

Construct validity

Quality of an evaluation method which faithfully reflects the construct, concept or theorised 'object' of evaluation. It must be expressed in sufficiently precise terms so that observations in the field allow for a reliable and sensitive analysis of the object being observed. For example, if the impact of support for innovation must be estimated by means of a survey of regional SMEs, the notion of innovation must have been defined precisely and in ways which properly represents the concept of innovation so that it can be measured and observed (e.g. number of new products or introduction of new production procedures).

Constructivism

A philosophical position within the social sciences which asserts that reality cannot be observed objectively but it is constructed through social interactions. It emphasises the responsive, interactive, dialogic and orchestrating role of the evaluator because the sources of data that are privileged are seen to reside with stakeholders and not with studies and externally generated data.

Context

The socio-economic environment in which an intervention is implemented.

Control group

A group consisting of eligible units which have been excluded from participation in the intervention by a process of random selection. Apart from its non-participation in the intervention, the control group is, from every point of view, comparable to the group of participants, and thus it represents the best approximation to the counterfactual (what would have happened to the treated had they not been exposed to the treatment).

Correlation

In statistics, correlation (often measured as a correlation coefficient) indicates the strength and direction of a linear relationship between two variables. That is in contrast with the usage of the term in colloquial speech, which denotes any relationship, not necessarily linear. A large degree of covariance between A and B indicates an assumption of causality but does not prove it. (Is A the cause of B? or is B the cause of A? or are A and B the consequence of something else?).

Cost-benefit analysis

Tool for judging the advantages of the intervention from the point of view of all the groups concerned, and on the basis of a monetary value attributed to all the positive and negative consequences of the intervention (which must be estimated separately). When it is neither relevant nor possible to use market prices to estimate a gain or a loss, a fictive price can be set in various ways. The first consists of estimating the willingness of beneficiaries to pay to obtain positive impacts or avoid negative impacts. The fictive price of goods or services can also be estimated by the loss of earnings in the absence of

those goods or services (e.g. in cases of massive unemployment, the fictive price of a day's unskilled work is very low). Finally, the fictive price can be decided on directly by the administrative officials concerned or the steering group. Cost-benefit analysis is used mainly for the ex ante evaluation of large projects.

Cost-effectiveness analysis

Evaluation tool for making a judgement in terms of effectiveness. This tool consists of relating the net effects of the intervention (which must be determined separately) to the financial inputs needed to produce those effects. The judgement criterion might, for example, be the cost per unit of impact produced (e.g. cost per job created). This unit cost is then compared to that of other interventions chosen as benchmarks.

Counterfactual situation

The situation which would have arisen had the intervention not taken place. The effect of an intervention is defined as the difference between factual and counterfactual situation. Since by definition we can never observe the counterfactual situation, we can never observe effects with certainty. Real world evaluation designs are based on an estimate of the counterfactual derived either from comparing subjects who were exposed to an intervention with a comparison group who were not exposed, or from examining subjects before and after exposure**.**

Coverage rate

Percentage of the eligible group which was actually affected by an intervention. The coverage rate is a result indicator which is important to quantify for monitoring purposes. The exposure rate is used to indicate the portion of the population targeted, which received information on the programme.

Credibility

Quality of the results and conclusions of an evaluation when they are logically supported by empirical facts and justified by an analysis of valid data. Credibility depends on several factors, including: reliability of data, soundness of the method, but also the reputation of the evaluator.

Criterion

Character, property or consequence of a public intervention on the basis of which a judgement will be formulated. For example, an employment incentive programme may be judged in terms of "costs per job created" or "percentage of support benefiting the long-term unemployed" (in the latter case, it is assumed that the higher the percentage, the better the intervention). An evaluation criterion should be explicit, that is, it must clearly show why the intervention will be judged better or worse. The types of criteria frequently used in evaluation are: performance, effectiveness, equity and sustainability. Thus, evaluation criteria may refer to different social values. To be used in an evaluation, a criterion should be accompanied by a norm (level of success at which an intervention will be considered good in terms of this criterion). An intervention is generally judged in terms of several criteria.

Cross sectional data

Data collected at a given point in time on several individuals. For example: comparative unemployment rates in European regions; comparative placement rates for women and men.

Deadweight

Change observed among direct beneficiaries following a public intervention that would have occurred even without the intervention. It is a special case of the counterfactual situation. The only plausible reason to distinguish between deadweight and counterfactual is that the former underscores the fact that resources have been used to fund activities that would have taken place even without public support.

Delphi Panel

Procedure for iterative and anonymous consultation of several experts, aimed at directing their opinions towards a common conclusion. The Delphi panel technique may be used in ex ante evaluation, for estimating the potential impacts of an intervention and later to consider evaluation findings.

Demand-side effect

The effect of an intervention which spreads through growing intermediate consumption of enterprises (supplier effect) and through the income generated within the region, and which, in turn, generates spending by households (multiplier effect).

Developmental evaluation

Evaluation processes, including asking questions and applying evaluation logic, to support program, project, product and/or organisational development. The evaluator is part of a team whose members collaborate to conceptualise, design and test new approaches in a long-term ongoing process of continuous improvement, adaptation and intentional change.

Differences in differences

Difference in differences is a non-experimental technique relatively simple to implement. A group on non-participants is selected which shares some characteristics with the intervention group. The outcome-variable for the two groups is measured before and after the programme. The control group is measured for "natural change" over the period, and this is compared with the change in the intervention group. The difference between the two changes gives an estimate for the impact of policy.

Direct effect

Effects of a public intervention on its direct beneficiary, excluding all repercussions on other groups.

Disparity

The fact that a region or group of people are in a situation which differs significantly from others. In general, the rationale of programmes financed by the EU Structural Funds consists of reducing socio-economic disparities between regions or social groups.

Displacement effect

Effect obtained in an eligible area at the expense of another area, or by a group of beneficiaries at the expense of another group within the same territory. Displacement effects may be intended (e.g. displacement of a public administration from the capital to a 'lagging' region) or unintended (the jobs created by a regional development programme resulted in the disappearance of jobs in other eligible regions).

Distributional effects

The differential effects of a policy reform, or the introduction of a new programme or policy on various groups / individuals. Distributional effects are often assessed ex-ante in order to anticipate impacts and ensure that policies do not bring unexpected results.

Diversification

Deployment of the production of a firm or a region beyond its traditional specialisation, with a view to creating new activities and/or setting up in new market segments.

Economic, social and territorial cohesion

In European Union policy it refers to reducing unacceptable gaps between regions, territorial categories (e.g. urban / rural) or social groups, from the point of view of their level of economic

development and social integration. Economic and social cohesion relates to disparities in regional income and quality of life, as well as to all differences between social groups' access to socio-economic resources including employment. It is the main objective of the European policy of the same name (see Article 130 A of the Treaty).

Effect

The (causal) effect of an intervention is taken as the difference between the outcome observed after an intervention has taken place, and the outcome that would have been observed had the intervention not taken place: the latter is defined as the "counterfactual". Thus, being the difference between something observed and something hypothetical, an effect can never be directly observed. An effect can be inferred, insofar as the evaluator finds a credible way to approximate what would have happened had the intervention not taken place.

Effectiveness

The term effectiveness has many possible meanings. The most common definition identifies effectiveness with "achievement of objectives". This leaves open the definition to the different meanings of "objectives". Objectives can be expressed quantitatively in terms of expected output or results. The effectiveness is evaluated simply by comparing what has been obtained with what had been planned: outputs and results indicators are all is needed. A more challenging definition of effectiveness relates it to the ability of a given action to produce a desired change: effectiveness is evaluated by comparing what is observed after the action has taken place with what would have happened without the action. For the latter type of evaluation, simple indicators are not sufficient; one needs data that allow recovery of the counterfactual situation and qualitative approaches to understand how the action has been implemented.

Efficiency

The term efficiency has many possible meanings. The most common definition identifies efficiency with obtaining a given output at the minimum cost or, equivalently, with maximizing output for a given level of resources.

Eligibility

The fact that a region, project or group of people has the required characteristics to benefit from an intervention or, more precisely, to receive assistance. Eligibility criteria follow directly from the rationale of the intervention.

Employability

An individual's ability to find or keep a job in a given socio-economic environment. Employability concerns the appropriateness of skills in relation to the requirements of the labour market. The employability of an individual can be examined indirectly on the basis of pre-established factors (e.g. qualifications, experience, mobility, existence of job offers).

Endogenous Development

Increase in economic activity based on internal competitive advantages within a region or territory. The main factors of endogenous development are human capital, entrepreneurial spirit, local savings and local innovation networks. By contrast, exogenous development concerns the inward transfer of capital, technology, know-how and skills.

Environmental Impact Assessment

Study of all the repercussions of an individual project on the natural environment. Environmental Impact Assessment is a requirement in the EU in the selection of major infrastructure projects. By contrast, Strategic Environmental Assessment refers to the evaluation of programmes and policy priorities. Environmental Impact Assessment consists of two steps: screening, which refers to an initial

overall analysis to determine the degree of environmental evaluation required before the implementation is approved; and scoping which determines which impacts must be evaluated in depth. The evaluation of environmental impacts examines expected and unexpected effects. The latter are often more numerous.

Equal opportunities

Mainly used to refer to equal access for women and men to employment, at the same level of remuneration and social advantages, in a given socio-economic context. Equal opportunities can be applied to characteristics other than gender, such as race, ethnicity, sexuality, etc.

Evaluability assessment

Technical part of pre-evaluation, which takes stock of available knowledge and assesses whether technical and institutional conditions are sufficient for reliable and credible answers to be given to the questions asked. Concretely, it consists of checking whether an evaluator using appropriate evaluation methods and techniques will be capable, in the time allowed and at a cost compatible with existing constraints, to answer evaluative questions with a strong probability of reaching useful conclusions. In some formulations it also includes an assessment of the likelihood of evaluation outputs being used.

Evaluation

The process of determining the merit or worth or value of something; or the product of that process. The special features of evaluation include a characteristic concern with cost, comparisons, needs, ethics, and its own political, ethical and cost dimensions; and with the supporting and making of sound value judgements, rather than hypothesis-testing. The term is sometimes used more narrowly (as is "science") to mean only systematic and objective evaluation, or only the work of people labelled "evaluators". (Scriven M., Evaluation Thesaurus)

Evaluation capacity

The institutional, human, resource, skill and procedural base for conducting evaluations in public policy and public management systems. This structural definition is embodied in expert evaluation units within governments or other public agencies and in commitment and practice that conducts evaluation and integrates this into decision-making and policy making. It is also sometimes understood in cultural terms: as a reflex to question, be open to criticism, to learn from practice and to be committed to using evaluation outputs.

Evaluation design

Technical part of the evaluation plan, the clarification of the links between evaluation questions, arrangements for data collection and analysis and how evaluative judgements will be made.

Evaluative question

Question that need to be answered by evaluators. There are three main types of evaluation questions: descriptive (what happened?), causal (to what extent are outcomes due to the intervention?) and normative (is the effect satisfactory?). An evaluation generally has several questions.

Evaluator

The people who perform the evaluation, usually in a team in complex programmes that require a mix of skills and competencies. Evaluators gather and interpret secondary data, collect primary data, carry out analyses and produces the evaluation report. They may be internal or external- vis-à-vis the commissioning body or programme managers. Evaluation teams may bring together a group of evaluators drawn from a single or several organisations (consortium).

Ex ante evaluation

Evaluation which is performed before programme implementation. For an intervention to be evaluated ex ante, it must be known with enough precision; in other words, a plan, at least, must exist. If the intervention still has to be planned from scratch, one would refer to a needs analysis rather than ex ante evaluation. This form of evaluation helps to ensure that an intervention is as relevant and coherent as possible. Its conclusions are meant to be integrated at the time decisions are made. It provides the relevant authorities with a prior assessment of whether development issues have been diagnosed correctly, whether the strategy and objectives proposed are relevant, whether there is incoherence between them or in relation to Community policies and guidelines, whether the expected impacts are realistic.

Ex post evaluation

Evaluation which judges an intervention when it is over. It aims to account for the use of resources, the achievement of expected (effectiveness) and unexpected effects (utility), and for the efficiency of interventions. It strives to understand the factors of success or failure, as well as the sustainability of impacts. It also tries to draw conclusions which can be generalized to other interventions.

Expert panel

A technique, similar to a survey, which relies on the necessarily subjective views of experts in a particular field. It is not recommended to rely on expert opinion as a sole data source, for example, because of problems with so-called "chatty bias" (Scriven M., Evaluation Thesaurus).

Explanatory theory

All the factors likely to explain changes observed following the public intervention. The scope of explanatory theory is far wider than that of the theory of action. Like the theory of action, it encompasses relations of cause and effect. It also covers any causes likely to explain the changes which have taken place, i.e. all confounding factors. Evaluation relies on a list of explanatory assumptions established with the help of experts, based on research and evaluation in similar fields.

External coherence

Correspondence between the objectives of an intervention and those of other public interventions which interact with it. If a national policy and an EU socio-economic programme are implemented in a complementary manner in the same territory for the purpose of developing SMEs, it can be said that there is external coherence.

External evaluation

An evaluation which is performed by persons outside the organisation responsible for the intervention itself.

External validity

Quality of an evaluation method which makes it possible to obtain conclusions that can be generalized to contexts (groups, areas, periods, etc.) other than that of the intervention being evaluated. Only strong external validity allows one to extrapolate from lessons learned during the implementation of the evaluated intervention.

Externality

Effect of a private action or public intervention which is spread outside the market. For example: a firm pollutes a river and causes an economic loss for a fish farm downstream; an engineer leaves the firm in which he or she was trained and applies his or her know-how in a new firm which he or she creates. By their very nature, externalities trigger private choices which cannot be optimised through the mechanisms of market competition. Only collective and often public decisions are able to promote

positive external effects and prevent negative ones. A large proportion of financial support allocated within the framework of European cohesion policy is aimed at promoting positive external effects which businesses do not seek to create themselves spontaneously.

Factor analysis

Factor analysis is a statistical method used to describe variability among observed variables in terms of fewer unobserved variables called factors. The observed variables are modelled as linear combinations of the factors, plus "error" terms. The information gained about the interdependencies can be used later to reduce the set of variables in a dataset. Factor analysis originated in psychometrics, and is used in behavioural sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data.

Feedback

Feedback is a process by which evaluation findings and results are communicated to interested parties. It can be used to shape or modify a programme and support learning in a formative or developmental evaluation. Feedback also refers to giving information to programme stakeholders and beneficiaries - and those who have cooperated with the evaluation by providing information and access.

Field of intervention

A set of interventions which are similar enough for their indicators to be harmonised and for comparisons to be made between different evaluations.

Focus group

Survey technique based on a small group discussion. Often used to enable participants to form an opinion on a subject with which they are not familiar. The technique makes use of the participants' interaction and creativity to enhance and consolidate the information collected. It is especially useful for analysing themes or domains which give rise to differences of opinion that have to be reconciled, or which concern complex questions that have to be explored in depth.

Forecast

Anticipation or prediction of likely future effect.

Formative evaluation

Evaluation which is intended to support programme actors, i.e., managers and direct protagonists, in order to help them improve their decisions and activities. It mainly applies to public interventions during their implementation (on-going, mid-term or intermediate evaluation). It focuses essentially on implementation procedures and their effectiveness and relevance.

Funding authority

Public institution which helps to finance an intervention. By extension, the term funding authority is also used for people who intervene on behalf of these institutions in the evaluation process: European Commission desk officers, officials from a national ministry; elected representatives from a regional or local authority.

Goals Achievement Matrix

The goals achievement matrix clearly sets out planned goals and marks them against objectives and the necessary steps / measures to achieve the goals. For example, Goal 1 could be to improve economic growth, which could have a number of policy objectives i.e. promote high value added economy, retain diverse economic structure and remove obstacles to intervention, which also have a number of measures / alternatives to achieving the objectives.

Gross effect

The term should be "gross change"-- that is, the change observed following a public intervention. Although the "gross effect" appears to be the consequence of the intervention, usually it cannot be entirely imputed to it. The following example shows that it is not sufficient for an evaluation merely to describe gross effects". Assisted firms have created 500 jobs utilizing a public subsidy (gross effect). In reality, they would in any case have created 100 jobs even without the support (deadweight). Thus, only 400 jobs are really imputable to the intervention (net effect). The challenge is to establish how many jobs would have been created without the public support - that is, to establish a counterfactual.

Impact

The change that can be credibly attributed to an intervention. Same as "effect" of intervention or "contribution to change".

Impartiality

Quality of conclusions and recommendations of an evaluation when they are justified by explicit judgement criteria and have not been influenced by personal or partisan considerations. An impartial evaluation takes into account the expectations, interpretations and judgement criteria of all legitimate stakeholders, including those who have very little power or ability to express themselves. Impartiality is an essential element of the quality of an evaluation.

Implementation

Implementation is the realization of an application, or execution of a plan, idea, model, design, specification, standard, algorithm, or a policy. For our purposes, implementation refers to the carrying out of public policy. Factors impacting implementation include the legislative intent, the administrative capacity of the implementing bureaucracy, interest group activity and opposition.

Impulsion effect

Secondary effect which spreads through investments induced upstream and downstream from the sector affected by the intervention. For example, the construction of a large infrastructure project generates the creation of new businesses in the region. These continue to expand after the work has ceased.

Income multiplier effect

Secondary effect resulting from increased income and consumption generated by the public intervention. Multiplier effects are cumulative and take into account the fact that part of the income generated is spent again and generates other income, and so on in several successive cycles. In each cycle, the multiplier effect diminishes due to purchases outside the territory. The effect decreases much faster when the territory is small and when its economy is open.

Independent

Separate and autonomous from the stakeholder groups involved in the intervention, and therefore able to provide impartiality.

Indicator

A characteristic or attribute which can be measured to assess an intervention in terms of its outputs or results. Output indicators are normally straightforward. Result indicators may be more difficult to derive, and it is often appropriate to rely on indirect indicators as proxies. Indicators can be either quantitative or qualitative. Context indicators relate to the environment for the programme.

Indirect beneficiary

A person, group of persons or organisation which has no direct contact with an intervention, but which is affected by it via direct beneficiaries (e.g. firms which have used technology transfer networks set up by a public intervention to innovate).

Indirect effect

Effect which spreads throughout the economy, society or environment, beyond the direct beneficiaries of the public intervention. Indirect "internal" effects, which are spread through market-based relations (e.g. effect on suppliers or on the employees of an assisted firm), are distinguished from external effects or "externalities" which are spread through non-market mechanisms (e.g. noise pollution; cross-fertilisation within an innovation network).

Individual interview

Technique used to collect qualitative data and the opinions of people who are concerned or potentially concerned by the intervention, its context, its implementation and its effects. Several types of individual interview exist, including informal conversations, semi-structured interviews and structured interviews. The latter is the most rigid approach and resembles a questionnaire survey. A semi-structured interview consists of eliciting a person's reactions to predetermined elements, without hindering his or her freedom to interpret and reformulate these elements.

Information system

Arrangements to store information on interventions, their context and progress (inputs, outputs and results) so that they can be accessed and inform decision makers, managers and evaluators. A monitoring information system may also include the syntheses and aggregations, periodically presented to the authorities responsible for the implementation (reviews, operating reports, indicators, etc.). In EU socio-economic programmes, the key element in an information system is a system of indicators. Information systems are a narrower concept than knowledge systems that combine records, lessons, syntheses and experience as well as routine data-sets.

Input

Financial, human, material, organisational and regulatory means mobilised for the implementation of an intervention. Monitoring and evaluation focus primarily on the inputs allocated by public authorities and used by operators to obtain outputs. This gives a relatively broad meaning to the word "input". Some prefer to limit its use to financial or budgetary resources only. In this case, the word "activity" can be applied to the implementation of human and organisational resources.

Input-output analysis

Tool which represents the interaction between sectors of a national or regional economy in the form of intermediate or final consumption. Input-output analysis serves to estimate the repercussions of a direct effect in the form of first round and then secondary effects throughout the economy.

Institutional capacity

The capacity of an institution to perform certain tasks and requirements.

Internal coherence

Correspondence between the different objectives of the same intervention. Internal coherence implies that there is a hierarchy of objectives, with those at the bottom logically contributing towards those above.

Internal validity

The kind of validity of an evaluation design that answers the question: "Does the design prove what it's supposed to prove about the treatment on the subject actually studied?". In particular, does it prove that the treatment produced the effect? Reasons for lack of internal validity include poor instruments, participant maturation, spontaneous change, or selection bias (Scriven M., Evaluation Thesaurus)

Intervention

Any action or operation carried out by public authorities regardless of its nature (policy, programme, measure or project). The term intervention is systematically used to designate the object of evaluation.

Learning

This can be a both a process and a product. As a product it refers to the fact that lessons drawn from experience are accepted and retained by institutions or organisations responsible for intervention. The learning goes beyond feedback insofar as the lessons are capitalised on and can be applied to other settings. As a process learning refers to the ways in which new data, information and experiences are accessed, internalised and accepted - as well as used. Lesson learning is widely recognised as a key output of evaluations ensuring that evaluation results are used and past mistakes not repeated.

Leverage effect

Propensity for public intervention to induce private spending among beneficiaries. In cases where public intervention subsidises private investments, leverage effects are proportional to the amount of private spending induced by the subsidy. Leverage effects must not be confused with additionality.

Logic models

Generic term that describes various representations of programmes linking their contexts, assumptions, inputs, intervention logics, implementation chains and results. These models can be relatively simple (such as the logical framework, see below) and more complex (such as realist, context/mechanism/outcome configurations and Theory of Change models).

Logical framework

Tool used to structure the logic of a public intervention. It is based on a matrix presentation of the intervention, which highlights its needs, objectives, inputs, outputs and results and other contributing factors.

Longitudinal data

Repeated observations of the same individuals at regular intervals, during a given period.

Mainstreaming

The term as applied to equal opportunities meant systematically taking into account the specific priorities and needs of women and men in all dimensions of an intervention, from the design and implementation stage to monitoring and evaluation.

Managerial evaluation

An evaluative approach integrated into the management of public interventions, and aimed at recommending changes related either to decision-making (feedback, instrumental use) or to the behaviour of the actors responsible for the implementation of the intervention. The general approach of managerial evaluation is similar to that of new public management, and is aimed at addressing the problem of stagnating public revenue. The underlying question can be formulated as follows: how can the trade-off between the different sectoral policies be justified? The dominant approach here, which occurs within the administration, is that of optimisation of budgetary resources.

Meta-evaluation

Evaluation of another evaluation or of a series of evaluations. Such syntheses, systematic reviews or meta analyses generally share the assumption that lessons are best learned cumulatively over more than one evaluation if one wants to have confidence in findings. Meta evaluations can focus on results, on the mechanisms that underpin different programmes and even on the contexts of programmes - especially when what is being synthesised is descriptive or narrative case studies. Results are often judged in terms of their reliability, credibility and utility. They can also be judged in terms of their generalization and likely sustainability.

Method

Methods are families of evaluation techniques and tools that fulfil different purposes. They usually consist of procedures and protocols that ensure systemisation and consistency in the way evaluations are undertaken. Methods may focus on the collection or analysis of information and data; may be quantitative or qualitative; and may attempt to describe, explain, predict or inform action. The choice of methods follows from the evaluation questions being asked and the mode of enquiry - causal, exploratory, normative etc. Understanding a broad range of methods ensures that evaluators will select suitable methods for different purposes.

Methodology

Most broadly, the overall way in which decisions are made to select methods based on different assumptions about what constitutes knowing (ontology) what constitutes knowledge (epistemology) and more narrowly how this can be operationalised, i.e. interpreted and analysed (methodology).

Mid-term evaluation

Evaluation which is performed towards the middle of the period of implementation of the intervention. Mid-term evaluation has a formative character: it provides feedback on interventions of which it helps to improve the management. Mid-term evaluation is a form of interim evaluation.

Monitoring

The continuous process of examining the context of the programme and the delivery of programme outputs to intended beneficiaries, which is carried out during the execution of a programme with the intention of immediately correcting any deviation from operational objectives.

Multicriteria analysis

Tool used to compare several interventions in relation to several criteria. Multicriteria analysis is used above all in the ex ante evaluation of major projects, for comparing between proposals. It can also be used in the ex post evaluation of an intervention, to compare the relative success of the different components of the intervention. Finally, it can be used to compare separate but similar interventions, for classification purposes. Multicriteria analysis may involve weighting, reflecting the relative importance attributed to each of the criteria. It may result in the formulation of a single judgement or synthetic classification, or in different classifications reflecting the stakeholders' different points of view. In the latter case, it is called multicriteria-multijudge analysis.

Need

Problem or difficulty affecting concerned groups or regions, which the public intervention aims to solve or overcome. Ex ante evaluation verifies whether the needs used to justify an intervention are genuine. Needs are the judgement reference of evaluations which use relevance and usefulness criteria.

Net effect

Effect imputable to the public intervention and to it alone, as opposed to gross changes or gross "effects". To evaluate net effects, based on gross changes, it is necessary to subtract the changes which would have occurred in the absence of the public intervention, and which are therefore not imputable to it (the counterfactual).

Nominal data

Data related to gender, race, religious affiliation, political affiliation, etc., are examples of nominal data. In a more general form the data assigned with labels or names are considered as the data in Nominal scale. Since, each label or name indicates a separate category in the data; this data is also called as categorical data. The only comparison that can be made between two categorical variables is that they are equal or not, these variables cannot be compared with respect to the order of the labels.

Norm

Level that the intervention has to reach to be judged successful, in terms of a given criterion. For example, the cost per job created was satisfactory compared to a national norm based on a sample of comparable interventions.

Normative aim

The values or assumptions that underpin a programme and its goals. Common examples might include: preventing the desertification of rural areas, increasing competitiveness, equal opportunities and sustainability.

Objective

Explicit statement on the results to be achieved by a public intervention. If the objectives are not stated explicitly evaluation (and particularly ex ante evaluation) may help to clarify them. A quantified objective is stated in the form of targets and a qualitative objective in the form of descriptors, e.g. 30% of all outputs must be produced by the end of the third year; the public intervention must first benefit the long-term unemployed

On-going evaluation

Evaluation which extends throughout the period of implementation of an intervention. This form of evaluation accompanies the monitoring of outputs and results. It is too often confused with monitoring. The advantage of on-going evaluation is that it allows for effective collaboration between the evaluator and programme managers, which in turn favours a better appropriation of conclusions and recommendations. On-going evaluation may be seen as a series of in-depth studies, comprising successive analyses of evaluative questions which have appeared during the implementation. In general on-going evaluations are formative in intent.

Ordinal data

Ordered categories (ranking) with no information about distance between each category. They are data where there is a logical ordering to the categories. A good example is the Likert scale: 1=Strongly disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly agree.

Output

An indicator describing the "physical" product of spending resources through policy interventions. Examples are: the length, width or quality of the roads built; the number of hours of extra teaching hours provided by the intervention; the capital investment made by using subsidies.

Participant observation

In situ, non-disruptive observation of the daily activity of actors and/or beneficiary of the evaluated intervention. The researcher tries to understand the situation "from the inside". Ethnographic observation is useful in little known situations or when access to the field is difficult. It is used to

collect very detailed information. It also serves to identify all the effects of the intervention and the influence of the context.

Participatory evaluation

Evaluative approach that encourages the active participation of beneficiaries and other stakeholders in an evaluation. They may participate in the design and agenda setting of an evaluation, conduct self evaluations, help gather data or help interpret results.

Partnership

Partnership can be defined as an arrangement whereby two or more parties co-operate and work together. Often the aim of the partnership is to co-ordinate the use of partners' resources more economically, efficiently and effectively. Generally, partners have a shared aim/set of objectives and develop a commitment to an agenda for joint or co-ordinated action. Ideally, partnerships should achieve synergy by pooling resources and co-operative action, avoiding duplication and achieving more together than each partner can achieve alone.

Peer review

The term to describe the process whereby peers (stakeholders of equivalent position / practice area) review policies or practice e.g. academic peer reviews occur whereby academics review each others' work / articles, peer reviews on labour market policy work.

Performance

The meaning of the word performance is not yet stable; it is therefore preferable to define it whenever it is used. Performance might mean that intended results were obtained at a reasonable cost, and/or that the beneficiaries are satisfied with them. Efficiency and performance are two similar notions, but the latter extends, more broadly, to include qualitative dimensions.

Performance Management

An approach to public management that focuses on results and how to achieve improved results within finite available resources.

Performance Reserve

A mechanism within the EU Structural Funds by which appropriations allocated to each Member State were placed in reserve until 2003. This was distributed to the best-performing programmes by 31 March 2004. The measure was designed to motivate the Fund recipients. It was replaced by an optional performance reserve in 2007-2013. Discussion is ongoing on whether there will be a performance reserve in 2014-2020.

Pluralistic Evaluation

Evaluative approach designed as a collective problem-solving process involving all the parties concerned. On the basis of reliable information acceptable to all, value judgements are formulated by seeking agreement within an evaluation authority consisting of political and administrative officials, as well as spokespersons for the groups concerned.

Policy

A policy is typically described as a plan of action to guide decisions and achieve rational outcome(s). The term may apply to government, private sector organizations and groups, and individuals. Policy differs from rules or law. While law can compel or prohibit behaviours (e.g. a law requiring the payment of taxes on income) policy merely guides actions toward those that are most likely to achieve a desired outcome.

Policy cycle

The policy cycle is the term used to describe the lifespan of a policy, from its formulation, to the review. It comprises: needs assessment / agenda setting; planning / policy formulation; policy implementation; policy monitoring; and evaluation and feedback.

Positivism

A belief that it is possible to obtain objective knowledge through observation and that such knowledge is verified by statements about the circumstances in which such knowledge is true. This objectivity is achieved by using objective instruments like tests or questionnaires.

Primary data

Data collected directly in the field, by means of a survey carried out by the evaluator on the groups concerned by the intervention. Primary data play an important role in the cognitive contribution of the evaluation. They are added to data already available at the start of the evaluation (e.g. former research and evaluations, monitoring data, statistics).

Programme

Organised set of financial, organisational and human interventions mobilised to achieve an objective or set of objectives in a given period. A programme is delimited in terms of a timescale and budget. Programme objectives are defined beforehand; an effort is then made systematically to strive for coherence among these objectives. The three main steps in the life-cycle of a programme are design, implementation and ex post evaluation review.

Programme cycle

The programme cycle follows the same pattern as the policy cycle. It contains the following stages agenda setting; planning / programme formulation; programme implementation; programme monitoring; and evaluation and feedback.

Project

Non divisible operation, delimited in terms of schedule and budget, and placed under the responsibility of an operator. For example: creation of a new training branch, extension of a purification network, carrying out of a series of missions by a consultancy firm. Within the framework of European Cohesion Policy, the operator requests assistance which, after a selection procedure, is either attributed or not by the managers of the programme. Particularly careful ex ante evaluations are made of major infrastructure projects, using the cost-benefit analysis technique.

Project promoter

Public or private person or organisation which requests and possibly obtains assistance in the framework of an intervention for a given project (e.g. rehabilitating a run down urban site; creating a new training branch). A project promoter should be considered to be an operator if it receives public funds every year and if it has to report regularly and permanently on the project. In contrast, it should be considered a beneficiary if it receives limited funding for a single project.

Propensity Score Matching

A statistical technique for constructing a control group. The process at its heart has the following steps: (1) identify key variables which are thought to predict membership in the treatment group, (2) use logistic regression to generate a scoring system, based on these variables, to predict the likelihood of belonging to the treatment group, (3) match each member of the treatment group with a control which has a similar score; (4) estimate the effect as a difference between means. PSM is an elegant and powerful process for generating a matching group where this might otherwise be difficult, but it is not a miracle cure.

Questionnaire survey

The basic instrument for surveys and structured interviews. Their design of them takes major skill and effort. Often too long, which reduces response rate as well as validity (because it encourages stereotyped, omitted, or superficial responses). They must be field-tested; usually a second field-test still uncovers problems e.g. of ambiguity. (Scriven M., Evaluation Thesaurus)

Ratio data

Ratio data is continuous data where both differences and ratios are interpretable. Ratio data has a natural zero. A good example is birth weight in kg.

Rationale

The fact that an intervention can be justified in relation to needs to satisfy or socio-economic problems to solve. Ex ante evaluation verifies the real existence of these needs and problems, and ensures that they cannot be met or solved by existing private or public initiatives. Thus, the inadequacy or shortcomings of other initiatives (whether private or public) may be a fundamental element in the programme rationale.

Realism

An approach to evaluation and research based on a philosophy of science that is concerned with 'real world' problems and phenomena but believes these cannot simply be observed. It seeks to open the black-box within programmes or policies to uncover the mechanisms that account for what brings about change. It does so by situating such mechanisms in contexts and attributing to contexts the key to what makes mechanisms work or not work. Different mechanisms come into play in different contexts which is why some programmes or policy instruments work in some but not all situations.

Regression analysis

Regression analysis refers to techniques for modelling and analysing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables — that is, the average value of the dependent variable when the independent variables are held fixed.

Relevance

Appropriateness of the explicit objectives of an intervention, with regard to the socio-economic problems the intervention is meant to solve. Questions of relevance are particularly important in ex ante evaluation because the focus is on the strategy chosen or its justification. Within the framework of mid-term evaluation, it is advisable to check whether the socio-economic context has evolved as expected and whether this evolution calls into question the relevance of a particular initial objective.

Reliability

Quality of the collection of evaluation data when the protocol used makes it possible to produce similar information during repeated observations in identical conditions. Reliability depends on compliance with the rules of sampling and tools used for the collection and recording of quantitative and qualitative information.

Reporting System

A reporting system can be any system through which information is recorded regarding a specific project or programme. As well as keeping records and useful information on progress, they assist in the monitoring process and the evaluation stages of any project or programme.

Result

The specific dimension of the well-being of people that motivates policy action, i.e., that is expected to be modified by the interventions designed and implemented by a policy. Examples are: the mobility in an area; the competence in a given sector of activity.

Result Indicator

An indicator describing a specific aspect of a result, a feature which can be measured. Examples are: the time needed to travel from W to Y at an average speed, as an aspect of mobility; the results of tests in a given topic, as an aspect of competence; the share of firms denied credit at any interest rate, as an aspect of banks' rationing.

Sample

In statistics, a sample is a subset of a population. Typically, the population is very large, making a census or a complete enumeration of all the values in the population impractical or impossible. The sample represents a subset of manageable size. Samples are collected and statistics are calculated from the samples so that one can make inferences or extrapolations from the sample to the population. The best way to avoid a biased or unrepresentative sample is to select a random sample, also known as a probability sample. A random sample is defined as a sample where the probability that any individual member from the population being selected as part of the sample is exactly the same as any other individual member of the population. Several types of random samples are simple random samples, systematic samples, stratified random samples, and cluster random samples.

Scope of evaluation

Definition of the evaluation object, of what is evaluated. The scope of the evaluation is usually defined in at least four respects: operational (all or part of the domains of intervention, one or several related policies), institutional (all or part of the authorities), temporal (period taken into consideration) and geographical (one or more territories or parts of territories, a particular region, town, nature reserve, etc.).

Scoring

Choice of a level on a scale graduated in quantitative measurement units (e.g. a scale of 0 to 100 or 3 to +3) in order to represent the significance of an effect, need or element of quality. It is possible to construct an observation grid which is sufficiently structured to directly produce a score. The person who chooses the score is called the scorer or the assessor.

Secondary data

Existing information, gathered and interpreted by the evaluator. Secondary data consists of information drawn from the monitoring system, produced by statistics institutes and provided by former research and evaluations.

Self-evaluation

Evaluation of a public intervention by groups, organisations or communities which participate directly in its implementation. Is usually complementary to other forms of expert or external evaluations.

Shift-share analysis

Tool for evaluating regional policy, which estimates the counterfactual situation by projecting national economic trends onto the economy of a given region. The basic assumption of this technique is that, in the absence of regional policy, the evolution of economic variables in the region would have been similar to that of the country as a whole. Comparison between the policy-off and policy-on situations

is concluded with an estimation of the macro-economic impact of regional policy. The optimum conditions for using this tool rarely exist.

Social partners

The organisations designated as representatives of both sides of industry in negotiations on pay and conditions, usually trade unions and employers organisations.

Socio-economic programmes

A programme that attempts to address both social and economic issues and bring social benefits alongside economic benefits.

Stakeholder

Individuals, groups or organisations with an interest in the evaluated intervention or in the evaluation itself, particularly: authorities who decided on and financed the intervention, managers, operators, and spokespersons of the publics concerned. These immediate or key stakeholders have interests which should be taken into account in an evaluation. They may also have purely private or special interests which are not legitimately part of the evaluation. The notion of stakeholders can be extended much more widely. For example, in the case of an intervention which subsidises the creation of new hotels, the stakeholders can include the funding authorities/managers, the new hoteliers (direct beneficiaries), other professionals in tourism, former hoteliers facing competition from the assisted hotels, tourists, nature conservation associations and building contractors.

Standard

A standard is a level of achievement along a normative dimension or scale that is regarded as the desired target to be achieved. Examples might include availability of childcare for all families with children under 6; air with a specified level of impurities; or populations with a certain qualifications profile.

Statistically significant

When the difference between two results is determined to be "statistically significant" one can conclude that the difference is probably not due to chance. The "level of significance" determines the degree of certainty or confidence with which we can rule out chance (rule out the "null hypothesis"). Unfortunately, if very large samples are used even tiny differences become statistically significant though they may have no social, educational or other value at all. (Scriven M., Evaluation Thesaurus)

Steering group

A steering group steers and guides an evaluation. It supports and provides feedback to evaluators, engages in dialogue in the course of the evaluation and is thereby better able to take seriously and use results. Steering committees may include the evaluation commissioner, programme managers and decision makers plus some or all of the other main stakeholders in an evaluated intervention.

Strategic Environmental Assessment

A similar technique to Environmental Impact Assessment but normally applied to policies, plans and programmes. Strategic Environmental Assessment provides the potential opportunity to avoid the preparation and implementation of inappropriate plans and programmes.

Strategy

Selection of priority actions according to the urgency of needs to be met, the gravity of problems to be solved, and the chances of actions envisaged being successful. In the formulation of a strategy, objectives are selected and graded, and their levels of ambition determined. Not all territories, sectors and groups are concerned by the same development strategy. Ex ante evaluation examines whether the strategy is suited to the context and its probable evolution.

Structural Funds

Structural Funds are the main financial instruments used by the European Union to reduce disparities and promote economic and social cohesion across European regions. Their combined efforts help to boost the EU's competitive position and, consequently, to improve the prospects of its citizens. The total budget for the Structural Funds amounts to 350 billion Euros in 2007-13, divided between the European Regional Development Fund (ERDF), the European Social Fund (ESF) and the Cohesion Fund.

Structuring effect

Structuring effects are changes which last after the public spending has ceased. They include sustainable effects at the micro-economic level and supply-side effects, but not demand-side effects. Structuring effects should not be confused with structural adjustment, which strives for the convergence of the macro-economic variables of a country towards international standards, particularly in terms of public finances and inflation.

Subsidiarity

In the European context, subsidiarity means, for example, that the European Union acts in those cases where an objective can be achieved better at the European level than at the level of Member States taken alone.

Substitution effect

Effect obtained in favour of a direct beneficiary but at the expense of a person or organisation that does not qualify for the intervention. For example, a person unemployed for a long time found a job owing to the intervention. In reality, this job was obtained because someone else was granted early retirement. If the objective was the redistribution of jobs in favour of disadvantaged groups, the effect can be considered positively.

Summative Evaluation

It is conducted after completion and for the benefit of some external audience or decision-maker (e.g. funding agency, historian, or future possible users). For reason of credibility it is much more likely to involve external evaluators than is a formative evaluation. (Scriven M., Evaluation Thesaurus)

Supply-side effect

Secondary effect which spreads through the increased competitiveness of businesses and thus of their production. The main mechanisms at play are increased productive capacity, increased productivity, reduced costs, and the diversification and reinforcement of other factors of competitiveness such as human capital, public facilities, the quality of public services, innovation networks, etc.

Sustainable development

Increase in economic activity which respects the environment and uses natural resources harmoniously so that future generations' capacity to meet their own needs is not compromised. By contrast, unsustainable development is characterised by the destruction of natural resources. This has negative repercussions on long-term development potential.

SWOT (Strengths, Weaknesses, Opportunities, Threats)

This is an evaluation tool which is used to check whether a public intervention is suited to its context. The tool helps structure debate on strategic orientations.

Synergy

The fact that several public interventions (or several components of an intervention) together produce an impact which is greater than the sum of the impacts they would produce alone (e.g. an intervention

which finances the extension of an airport which, in turn, helps to fill tourist facilities, also financed by the intervention). Synergy generally refers to positive impacts. However, phenomena which reinforce negative effects, negative synergy or anti-synergy may also be referred to (e.g. an intervention subsidises the diversification of enterprises while a regional policy helps to strengthen the dominant activity).

Target group

The intended beneficiaries (individuals, households, groups, firms) of an intervention. An intervention may have more than one target group. This term should be distinguished from "population" in the statistical sense.

Terms of reference

The terms of reference define the work and the schedule that must be carried out by the evaluation team. It normally specifies the scope of an evaluation, states the main motives and the questions asked. It sums up available knowledge and outlines an evaluation method - although offering scope for innovation by proposers. It describes the distribution of the work and responsibilities among the people participating in an evaluation process. It fixes the schedule and, if possible, the budget. It specifies the qualifications required from candidate teams as well as the criteria to be used to select an evaluation team.

Thematic evaluation

Evaluation which transversally analyses a particular point (a theme) in the context of several interventions within a single programme or of several programmes implemented in different countries or regions. The theme may correspond to an expected impact (e.g. competitiveness of SMEs) or to a field of interventions (e.g. R&D). The notion of thematic evaluation is similar to that of an in-depth study (e.g. impact of support for R&D on the competitiveness of SMEs), but it is a large scale exercise when conducted on a European scale.

Theory of action

All the assumptions made by funding authorities and managers to explain how a public intervention will produce its effects and achieve its aim. The theory of action consists of relations of cause and effect linking outputs, results and impacts. It is often implicit, or at least partly so. Evaluation helps to clarify the theory and for that purpose often relies on various forms of programme theory or logic models.

Time series

Data collected on the same population, in a comparative way, at regular intervals during a given period. Statistics institutes and statistical teams are the main sources of time series data.

Tool

Standardised procedure which specifically operationalises a method. A method might be gathering the views of SME managers; a tool might be a survey; and a technique might include a self completion questionnaire using 5 point scales.

Unit of Analysis

The unit of analysis is the unit that is being analysed in an evaluation. For instance, any of the following could be a unit of analysis in a study: individuals, groups, artefacts (books, photos, newspapers), geographical units (town, census tract, state), social interactions (divorces, arrests).

Utility

The fact that the impacts obtained by an intervention correspond to society's needs and to the socio-economic problems to be solved. Utility is a very particular evaluation criterion because it disregards all reference to stated objectives of an intervention. It may be judicious to apply this criterion when objectives are badly defined or when there are many unexpected effects. The criterion must, however,

be used with caution to avoid the evaluation team being influenced by personal considerations in their selection of important socio-economic needs or problems. Some authors have argued for a form of goal-free evaluation.

Value for money

Term referring to judgement on whether sufficient impact is being achieved for the money spent. It is often calculated by dividing the total project costs by the number of beneficiaries reached, and comparing the cost with alternative comparable measures in relation to the target groups and desired impacts.

Variance

A measure of how spread out a distribution of scores is. It is computed as the average of the squared deviations of each score from the mean.

Verifiable objective

An objective stated in such as way that it will subsequently be possible to check whether or not it has been achieved. A way of making an objective verifiable is to quantify it by means of an indicator linked to two values (baseline and expected situation). An objective may also be verifiable if it is linked to a descriptor, i.e. a clear and precise qualitative statement on the expected effect.

Weighting

A procedure to specify if one criterion is more or less importance than another in the formulation of a global judgement on an intervention.

Welfare

Welfare can be either a person's state of well-being, or it can refer to the social protection system, i.e. education, health provision.

Welfare economics

The branch of economics dealing with how well off people are, or feel that they are, under varying circumstances. It is sometimes regarded as the normative branch of economics. One application of welfare economics is cost benefit analysis, which attempts to balance the gains and losses from a proposed policy.