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

Table of Contents

WELCOME **1**

INTRODUCTION **3**

EVALUATING SOCIO-ECONOMIC DEVELOPMENT 3
WHO IS THIS GUIDE FOR? 3
WHY ANOTHER EVALUATION GUIDE? 3
UPDATING MEANS 4
CONTENT 5

CHAPTER 1: EVALUATION AND SOCIO-ECONOMIC DEVELOPMENT **7**

THE BENEFITS OF EVALUATION 7
1) Evaluations that make a difference *7*
2) Improving policies over time *8*
3) Designing programmes *9*
4) Choosing between instruments *9*
5) Improving management and delivery *9*
6) Identifying outputs, results and impacts *10*
7) Identifying unintended consequences and perverse effects *10*
8) Levels of evaluation: policies, themes, programmes and projects *11*
HISTORY AND PURPOSE OF EVALUATION 12
1) A short history of evaluation *12*
2) Different traditions and sources *16*
3) The main purposes of evaluation *17*
4) Learning as an overriding evaluation purpose *18*
METHODS AND THEIR ROOTS 18
1) Methodological axes *19*
2) Three philosophical traditions *20*
3) Four types of theory relevant to evaluation *23*
EVALUATION TO STRENGTHEN SOCIO-ECONOMIC DEVELOPMENT 24
1) How we evaluate depends on what we evaluate! *24*
2) The EU policy context *27*
3) Theoretical underpinnings of socio-economic development *29*
4) Community added value *31*
5) Doing the best in an imperfect world *32*
GOLDEN RULES 33

CHAPTER 2: DESIGNING AND IMPLEMENTING EVALUATION **35**

DESIGNING AND PLANNING YOUR EVALUATION 35
1) Evaluation and programming *35*
2) Planning evaluation *37*
3) Defining evaluation questions and criteria *40*
4) Choosing evaluation methods, responsibilities and allocating resources *45*
5) Writing the Terms of Reference *48*
IMPLEMENTING AND MANAGING EVALUATIONS 52
1) Choosing the right evaluators *52*
2) Managing the evaluation process *54*
3) Managing evaluation communications *59*
4) Managing quality assurance and quality control *61*
THE USE OF EVALUATION AND KNOWLEDGE MANAGEMENT 72
GOLDEN RULES 73

CHAPTER 3: DEVELOPING CAPACITY FOR SOCIO-ECONOMIC EVALUATIONS **75**

INTRODUCTION 75
EVALUATION CAPACITY IS PART OF INSTITUTIONAL CAPACITY 75
WHAT IS EVALUATION CAPACITY? 76
THE "ARCHITECTURE" OF EVALUATION SYSTEMS 79
STRENGTHENING EVALUATION DEMAND 80

1) Governments as a source of demand for evaluation ... 80
2) Civil society as a source of demand for evaluation ... 82
SUPPLY SIDE ... 83
1) Starting Conditions and Traditions ... 83
2) Links with the education, training and university sector ... 84
3) Independent Consultants: encouraging new suppliers ... 85
4) A professional evaluation community ... 86
INSTITUTIONALISING EVALUATION ... 88
GOLDEN RULES ... 93

CHAPTER 4: CHOOSING METHODS AND TECHNIQUES ... 95

FACTORS INFLUENCING THE CHOICE ... 96
1) Choosing methods and techniques ... 96
2) Quantitative versus Qualitative : A false debate? ... 98
3) Obtaining and using data and evidence ... 100
CHOICES FOR DIFFERENT TYPES OF INTERVENTIONS ... 101
1) Policy and Sectoral priorities ... 101
2) Local and territorial development ... 101
CHOICES FOR DIFFERENT EVALUATION PURPOSES ... 103
CHOICES FOR DIFFERENT PROGRAMME/POLICY STAGES ... 105
1) Formulation: Identifying needs and priorities ... 105
2) Design: Interventions and organisation ... 106
3) Implementation: Feedback and intermediate outcomes ... 106
4) Conclusions: Results and impacts ... 107
ACQUIRING AND USING DATA IN EVALUATION ... 107
1) All data are "produced" ... 107
2) Accessing data as a planned activity ... 108
3) Quantitative and Qualitative data ... 110
CREATING INDICATORS AND INDICATOR SYSTEMS ... 111
1) What are indicators? ... 111
2) Indicators and evaluation ... 114
3) Selecting indicators ... 116
4) Assessing quality ... 118
5) Applying indicators ... 120
USING INDICATORS TO IMPROVE MANAGEMENT ... 121
1) Managing performance rather than resources ... 121
2) Interpreting and comparing indicators ... 122
3) Avoiding the adverse effects of indicators ... 123
4) The usefulness of a system of indicators ... 125
GOLDEN RULES ... 126

ANNEX A - THE MAIN STAGES OF EVALUATION ... 129

EX-ANTE EVALUATION ... 129
SUCCESSIVE PROGRAMMING CYCLES ... 131

GUIDE ANNEX B - CHANGES IN STRUCTURAL FUND REGULATIONS ... 133

1988-1993 ... 133
1994-1999 ... 133
2000-2006 ... 134
2007-2013 ... 134

BIBLIOGRAPHY ... 137

GLOSSARY A – Z ... 139

Welcome

Evalsed[1] is an online resource providing guidance on the evaluation of socio-economic development. It promotes exchange of experience and good practice as well as increasing the relevance of evaluation in decision-making. Whilst the resource 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 main parts: THE GUIDE and three SOURCEBOOKS

The GUIDE is designed primarily for decision-makers - specifically those who design and manage evaluations in order 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 methods as well as providing guidance on how to design and implement evaluations and how to ensure their quality.

The SOURCEBOOKS (available only online) are of particular interest to practitioners and those wishing to impart or acquire evaluation skills. The three SOURCEBOOKS cover the following issues: evaluation approaches for particular themes and policy areas, evaluation methods and techniques and evaluation capacity building. The material in the Sourcebooks is updated regularly.

This printed publication concerns the Guide, the Bibliography and the Glossary.

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

[1] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/index_en.htm

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. The programmes are one means of 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 measure/intervention/project or at the level of programme evaluation. Many different programmes and their elements contribute to thematic objectives and thematic evaluation is another approach to evaluation. 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 thematic, policy and project evaluations as well as programme evaluation.

Who is this GUIDE for?

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

- Policy makers who have an interest in what evaluation can do for them including its strengths and limitations and the resources and capacities they will need,
- 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, we have also taken account of more specialist needs by preparing a number of sourcebooks to back up the content of the GUIDE.

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[2] - on indicators, ongoing evaluation, etc. - that is specific and closely aligned with the Structural Fund Regulations. There is also

[2] http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm

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 the socio economic development programmes 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 that need to be informed. We need to be clear that this Guide is not a substitute for other sources and indeed it draws on and cross-refers where relevant to such sources. This Guide 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 has 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 Agenda[3]. The role of human and social capital, the importance of information society and the knowledge economy as a means of achieving greater competitiveness and priorities of sustainable development and equal opportunities have all been brought to the fore.
- The accession of ten new member states in 2004 and two more in 2007 also poses 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. The authors of this GUIDE have had in mind throughout its preparation, the capacity building needs of many new member states. In practical terms we are concerned to maximize what can be achieved pragmatically with available resources, skills, institutional arrangements and data sources.
- 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 introduces both a more strategic and flexible approach, requiring more emphasis on specific evaluation questions.

Together these changes are substantial and this GUIDE was written and has been updated to take account of their consequences. It has also been planned and written to anticipate future change. There is no reason to believe that the future pace of change of evaluation and of socio economic policy will slow down. For this reason and to enable the ready updating of material in the GUIDE, Sourcebook material and the Glossary are accessible and searchable via the Internet.

[3] http://ec.europa.eu/growthandjobs/index_en.htm

Content

The new Evalsed GUIDE is published as single volume. It is supported by a series of Sourcebooks that provide more specialised and in depth material and which can be accessed and downloaded via the internet.

The Evalsed GUIDE itself is in four sections:

Evaluation and socio economic development provides an introduction to evaluation and its benefits. This begins with a general overview of what evaluation can do to improve policies and programmes and ultimately to strengthen socio-economic development. This is followed by an introduction to some basic ideas in evaluation: its history; some of the different traditions which evaluators draw on; and the different purposes of evaluation. Finally, the specifics of socio-economic development as an object of evaluation are discussed. This includes unpicking the specific characteristics of this socio economic development policy and its implications for evaluation as well as the main theories and ideas on which policies and programmes are built and which evaluators need to take into account.

Designing and implementing evaluation takes readers through practical issues in designing and implementing evaluations. It begins by considering design issues including how to plan an evaluation, defining evaluation questions and choosing methods, as well as launching and commissioning evaluations. It then goes on to consider the management issues once an evaluation has been designed including the choice of evaluators and the role of Steering Committees, managing quality assurance in evaluation and managing communications to ensure influence.

Developing evaluation capacity discusses how to develop evaluation capacity and strategies for capacity development are discussed. The argument is structured on the basis of four stages in capacity development. This part includes discussion of internal capacity within administrations, as well as external capacity within professional networks and partnerships.

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 of the Evalsed GUIDE 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 in the future is a theme that runs through this GUIDE.

To support the Evalsed GUIDE a series of Sourcebooks[4] has also been prepared, which will be of particular interest to specialists and practitioners.

[4] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/index_en.htm

Sourcebook 1 is entitled 'Evaluation approaches for particular themes and policy areas'. It considers important priorities such as sustainable development, the information society, social inclusion and equal opportunities and the range of policy areas within which interventions to promote socio economic development take place. The types of evaluation methods, data and indicators that are appropriate to these themes and policy areas are elaborated and examples of their application provided.

Sourcebook 2 is entitled 'Evaluation methods and techniques'. This includes the elaboration of a wide range of tools and techniques both quantitative and qualitative that are useful at different stages of an evaluation.

Sourcebook 3 is entitled 'Resource material on evaluation capacity building'. This includes case histories of the development of evaluation capacity in the EU, Italy, Netherlands and Ireland; and, references to other regional, national and international experience. It illustrates the advice provided in the Guide and is intended to stimulate the development of evaluation capacity.

There is a Glossary that contains definitions of the terms used in the Evalsed Guide and Sourcebooks.

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 programme has its own more specific rationale. Some may emphasise the development of modern transport and environmental infrastructure, some the regeneration of inner cities, some the building of basic transport and environmental infrastructure, some the modernisation of obsolete or declining industrial sectors, 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 procedures 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 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, as will be discussed further in this GUIDE, is certainly complex and often faces many uncertainties: it is not a precise science. Choosing goals and measures, designing programmes, 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 and usable, 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/programme cycle. This is why this GUIDE gives so much attention to the design of evaluation systems and the development of evaluation capacity inside public agencies and within professional and knowledge networks.

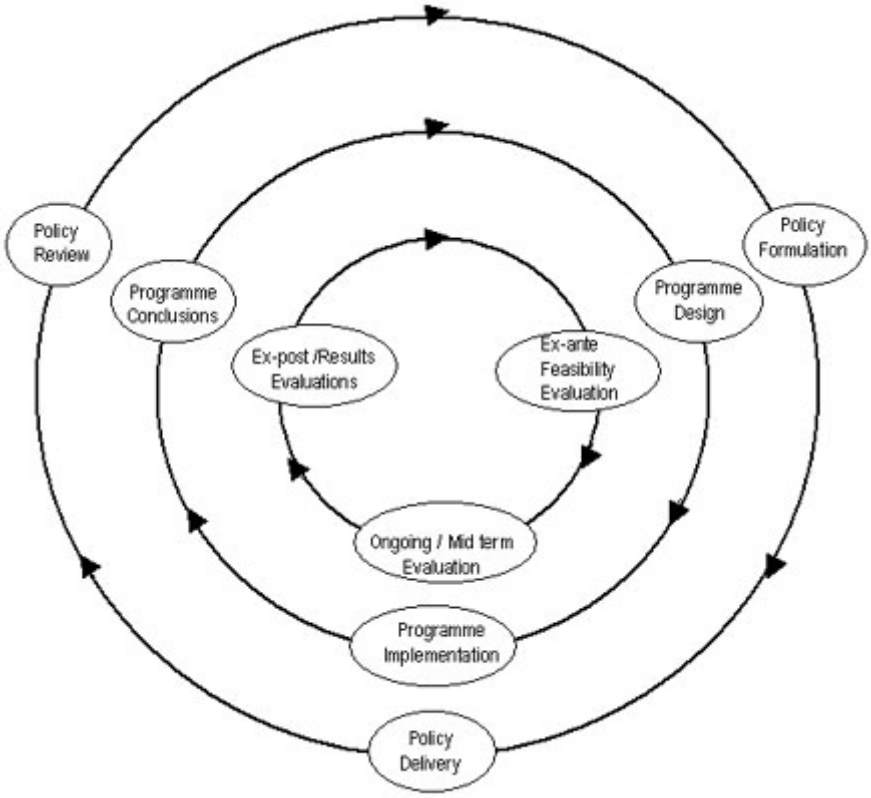
Second, evaluators and those who commission and use evaluation findings always need to balance best available methods with the demands of pragmatism. In the real world of socio-economic development we rarely have the time or resources - or even the data - to implement a comprehensive State of the Art evaluation. This is why this GUIDE places such 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 programme outputs and results. Evaluation language often follows this cycle as we can see from terms such as ex ante, mid-term and ex post evaluation commonly used in European Cohesion Policy. These terms are elaborated in Annex A. A similar logic is present in the distinction often made between outputs, results and impacts.

Before returning to the specific characteristics of socio-economic development, the contribution that evaluation can make at each stage of the policy cycle is first described. As the diagram in Policy, programme and evaluation cycles suggests, 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.

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 diagram does illustrate some of the main timing problems familiar in evaluation. The inner circle moves from ex ante evaluation that documents starting needs and the feasibility of planned programmes, through to ongoing or mid term evaluation that documents progress and implementation and finally to ex post evaluation that focuses on results. However, ex

ante evaluations should feed into programme design and to policy formulation, just as mid-term or interim evaluations should help shape programme implementation and policy about delivery of this and similar programmes. At the end of the evaluation cycle, ex post evaluations should contribute to policy reviews. Getting these cycles to align is desirable but does not always happen. 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. 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 to measure results and 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 all of these linked activities.

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 the needs of users is essential at this stage and evaluation can contribute to this. This input on relevance is not confined to programme design at one point in time. 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. Indeed an explicit re-programming moment is common.

4) Choosing between instruments

A well-designed evaluation system and particularly ex ante evaluation will also contribute to the selection of specific instruments or interventions within the general scope of a programme. Evaluation can make various inputs into selecting instruments. 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. Alternatively there may be 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 logic-model. 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, results and impacts

From an early stage, socio-economic development programmes need to demonstrate results. 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. Such results are expected to have a bigger impact that relate to policy and programme goals in the longer term. This allows evaluation to 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, identifying, describing and quantifying such outputs, results and impacts is a major benefit of evaluation. However, to make this process really useful policy makers need to ensure 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 monitoring and indicator systems are in place and that there are clear links between the indicators chosen and underlying objectives and goals.

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 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 and projects

The problem of linking policies, programmes 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 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. For example, if a well-structured framework has been designed for project self-evaluation there can be some assurance that the outcomes will be systematic and would merit further analysis at programme level. In addition, the 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 work need to be clear of the links between policy, programme, 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 post-war years in North America. Three strands that were most important in that early period were the evaluation of educational innovations (e.g., the effectiveness of new curricula in schools); linking evaluation with resource allocation (e.g., through a Planning, Programming and Budgeting system); and the evaluation of anti-poverty programmes (e.g., the Great Society experiments of the 1960s). 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 qualitative indicator 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, eg. teachers in education or scientists in research;
- managers and administrators, e.g., civil servants and managers of local 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, in particular, that had close links with the United States and Canada. From the 1970s onwards evaluation began to take root in different 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 has, until recently, 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 or furthering the practice of evaluation throughout the EU. At every stage of the programming cycle (ex-ante, mid-term or ongoing, 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. From modest beginnings in 1988, there is now an elaborated Structural Fund evaluation approach[5]

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, mid-term or ongoing, 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.

Over recent years there has been an evolution in Structural Fund regulations concerning evaluation.

The Evolution of Structural Fund Regulations:

1994 - 1999	2000 - 2006	2007-2013
EX-ANTE EVALUATION		
An ex-ante evaluation must be carried out in partnership by the Commission and the Member State; it must include the environmental impact; The Commission assesses the plans taking into account the ex-ante evaluation	The Member State has primary responsibility for the ex-ante evaluation. The aim of the ex-ante evaluation is defined; special attention must be given to the impact on the environment, on the labour market and on equality between the sexes. The ex-ante evaluation is incorporated in the plans	The Member State has primary responsibility for the ex ante evaluation. The aim of the evaluation is defined and the results are incorporated in the programmes. Proportionality was introduced in that ex ante evaluation was not obligatory for the National Strategic Reference Framework, while for small Regional Competitiveness and Employment Objective programmes, Member States could choose to evaluate programmes together.

[5] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/annex_b/reg_1988_1993_en.htm

MID-TERM EVALUATION		
There is no stipulation requiring information to be collected and communicated; There are no provisions defining an evaluative role for the Monitoring Committee (in practice, the main role of the Monitoring Committees has been to appoint evaluators).	The Member State is responsible for the mid-term evaluation in partnership with the Commission; the Commission assesses the evaluation's relevance; The objective of the mid-term evaluation is defined. The evaluation is organised by the relevant programme's managing authority and is carried out by an independent evaluator; it must be completed by 31 December 2003.To facilitate the various evaluations, the managing authority must create a system for collecting the financial and statistical data needed for the mid-term and ex-post evaluations, and must provide the information required for the ex-post evaluation. The Monitoring Committee examines the findings of the mid-term evaluation and may propose changes to the programme on that basis. An update of the mid-term evaluation is carried out by the end of 2005 by way of preparing the ground for operations thereafter. (The update of the Mid Term evaluation is also known as the final evaluation).	The Member State is responsible for ongoing evaluation, in consultation with the Commission. The 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 obligatory when there is a significant departure from the goals originally set.
EX-POST EVALUATION		
An ex-post evaluation must be carried out in partnership by the Commission and the Member State, assessing the impact of the measures taken in terms of the intended objectives.	The Commission has primary 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.

PERFORMANCE RESERVE		
	Before 31 March 2004, 4% of the indicative allocation for each Member State held back at the beginning of the period is allocated as a 'performance reserve' to the programming documents whose performance the Commission, on the basis of proposals from the Member State, considers to be successful. Performance is assessed on the basis of indicators of effectiveness, management and financial implementation which are defined by the Member State in consultation with the Commission.	The performance reserve is optional for Member States.

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.

These changes have been accompanied by shifts in responsibility between the different actors at European, national and regional levels and with the extension of the partnership principe. The range of potential stakeholders in evaluation has therefore expanded to include, for example, local authorities, social partners and civil society groups.

The reform of the Structural Funds Regulations (see Annex B) for the third generation of programmes (2000-2006) whilst devolving many obligations for evaluation to responsible authorities in Member States, required that these evaluations were used both at the ex-ante stage and again at the mid-term. This combination of devolved responsibilities with external scrutiny by higher tiers of government is also typical of national evaluations of socio-economic development.

Based on the experience of the 2000-2006 period, the major innovations of the 2007-2013 evaluation provisions are 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.

In recent years there has also been a strong move towards public management reform and the introduction of performance management concepts in many European countries as well as in the European Commission itself .

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 of evaluation, 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 in particular curriculum innovation in schools; and budgetary control and efficiency systems such as Planning, Programming and Budgetary Systems (PPBS). In reality these are only three particular 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 has begun to feature more prominently in evaluation in recent years 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. In addition, contemporary political perspectives highlight the importance of consensus building in order to strengthen legitimacy of policy action.

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.
- 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. It is these stakeholders who want to know whether the programmes assumptions are being 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 even 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 evaluation from this perspective has as its purpose:

> *to learn through systematic enquiry how to better design, implement and deliver public 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 positions 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 categories or families of method that will be discussed in Chapter 4 and are introduced here.

1) Methodological axes

In terms of methodologies, looking across the different approaches to evaluation discussed above, we can distinguish five methodological positions:

- The resource allocation position, 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 position, which is concerned to judge success and performance by the application of criteria.
- The explanatory position, which is concerned to explain programme impacts and success and make causal statements about what works, when and how.
- The formative or change oriented position, which provides positive and more complex feedback to support monitoring self correction during the life of a programme.
- The participatory/development position, 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 position.
- An indicator study that attempts to assess whether a programme has met its objectives would be an example of the *standards or target* based position.
- A mid-term evaluation or ongoing 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* position.
- 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* position.
- A locally led and focused 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 position.

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

Each of these policy areas or domains brings with it its own particular 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, goals and policy expectations in different policy areas which affect the kinds of measures or 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. It would be usual to commission an economic appraisal of infrastructure projects prior to the commitment of resources. This would be less common for projects where infrastructure investments were a smaller part of the total inputs being made in support of a particular intervention.

2) Three philosophical traditions

Three philosophical traditions underpin the broad methodological approaches to evaluation that are 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 which when analysed by objective techniques should lead to the same results whoever applied the technique. 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 quantifiable evidence on the relationships between the inputs of interventions and their outcomes.

The limitations of the tradition in the context of the evaluation of socio economic development 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.

The limitations of a pure form of positivism are now well recognised. Among these 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. To elaborate:

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.

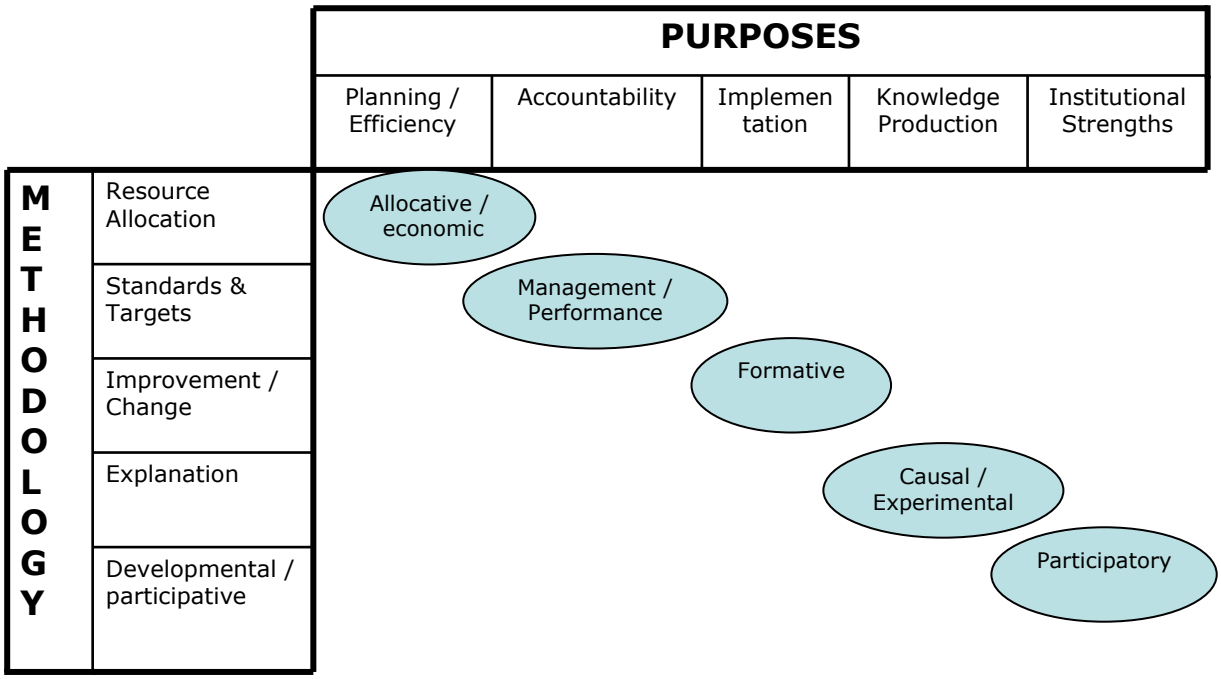
Realism seeks to open up the black box within policies and programmes to uncover the mechanisms that account for change. In doing so the tradition recognises that programmes and policies are embedded in multi layered social and organisational processes and that account should be taken of the influence 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. The evaluator is likely to form teacher learner relationships with policy makers, practitioners and participants. Thematic evaluations tend to operate within this tradition. 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 in *The two axes of evaluation, Purpose and Methodology*. Of course specific evaluation assignments will normally have more than one purpose and will apply more than one type of methodology. For example, a formative evaluation intended to improve programme implementation can also contribute to knowledge production, just as performance management types of evaluation can help strengthen institutions and networks if conducted appropriately. Thus a particular evaluation will not necessarily fit neatly within one of the five types. Rather it is likely to approximate to type. At the same time an evaluation system is likely to a reflect elements of several types of evaluation.

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 are:

Allocative / Economic evaluations at the intersection of planning and efficiency purposes and resource allocation methodologies;

Management / Performance oriented evaluations, at the intersection of accountability as a purpose and standard setting and targets as a methodology;

Formative evaluations, at the intersection of implementation and delivery as a purpose and methodologies aimed at improvement and change;

Causal / Experimental evaluations at the intersection of knowledge production as a purpose and methodologies that seek to offer explanations;

Participatory evaluations at the intersection of institutional (or network) strengthening as a purpose and social/organisational development methodologies.

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 purposes 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. We have, for example, noted the importance of performance management, of participation and

strengthening capacity; and all of these are incorporated in the types of evaluation identified in this framework.

These five 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 accompanying Sourcebooks.

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 the effectiveness of policies and programmes and evaluations that aim to advise, develop and improve policies and programmes. The ex post evaluation of the European Cohesion Policy falls into the former category while there is an expectation that ex ante and mid-term or ongoing evaluations will contribute more to programme improvements.

This GUIDE attempts to span both broad approaches even though they have different implications. For example external evaluations that aim to assess and measure effectiveness are more concerned with the structures of evaluation, including their independence. On the other hand, internal evaluations are less concerned with independence than in offering timely advice and inputs that can strengthen programmes from within. Users of this GUIDE need to bear this distinction in mind.

3) Four 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 four most common are:

- Programme Theory
- Theories about evaluation
- Theories of implementation and change
- Policy specific theory

Programme theory : the elaboration of the 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 that is concerned with opening up the black box and going beyond input output descriptions and seeking to understand what are the theories of actors with regard to programme interventions and why they should work.

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.

Policy specific theories: There is a body of theory associated with socio-economic development, i.e., how does development occur spatially and sectorally. There are similar bodies of theory in most policy priority areas, e.g., in education; health; employment; environmental protection, etc.. Sourcebook 1 elaborates on the theories linked to each of the themes and policy areas.

Because the design of interventions is usually underpinned by a rationale that derives from theory within policy areas, it is both useful and normal that evaluators have some knowledge of the theories relevant to the themes and policy areas under consideration.

Evaluation to strengthen socio-economic development

1) How we evaluate depends on what we evaluate!

This is not the place to analyse in all its diversity the nature of socio-economic development. However some introductory remarks are needed because how we evaluate is closely related to what it is that is being evaluated. Evaluators speak of the need to be clear about the object of evaluation and socio-economic development is a very particular object.

Of course socio-economic development encompasses many possible interventions including 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 of socio-economic development.

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 and Cohesion Policy interventions are certainly closely linked with regional policy. This dimension is stronger in some programmes and priorities than others (e.g., local development programmes in urban and rural areas) but always present;
- 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 both about growth in numbers (of income, jobs, firms, etc.) and also about 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. For example, the fact that the Lisbon Strategy seeks both to increase employment as well as achieve a dynamic knowledge-based economy expresses certain value priorities.

This has consequences for evaluation. For example:

- Methods adopted 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 essential requirement of socio-economic development evaluation;
- Ex ante assessment, pre-programming or planning evaluations need to identify resources on which development can build;
- Alongside quantitative measures (descriptions, indicators and models) qualitative evidence of the content, style standards and relevance of measures also need to be assessed;
- Policy frameworks and associated normative or value statements (e.g., about social solidarity, sustainable development, equal opportunities) will define key criteria for an evaluation and the evaluation focus.

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 industries are in declining sectors, firms are not competitive, public administrations have limited capacities, social groups now excluded have been excluded for a long time, education and training systems are poorly linked to economic needs and the infrastructure is poor. These circumstances can be compared with programmes where interventions are less ambitious, for example, where they involve introducing a new qualification system or upgrading an existing road system.
- 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 specific sectoral or thematic programme is likely to include multiple interventions or measures. This follows from an understanding that many persistent, 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 in comparison with many public interventions that are standardised across all 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 socio-economic cohesion across the EU. This would also be true of many national and regional socio-economic programmes.
- They have a strong bottom-up as well as a top down quality. 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. In many socio-economic programmes that adopt a local development approach, such partners take the lead in setting priorities, designing programmes and implementing and managing programme outputs.

- Being committed to structural and systemic change, socio-economic development programmes may not always achieve their long-term ambitions within a single programming period. Furthermore, because they have long term ambitions, such policies and programmes are usually concerned with the sustainability and viability 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 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.

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 and measures are assumed to interact and reinforce each other	Evaluation must analyse the interaction between many programmes/interventions. Evaluation should consider complexity and synergy
Programmes matched to settings	Programmes and measures 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 are important	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 methods. Where local/regional actors are dominant (e.g., territorial/ local development) evaluation should support self management/direction
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

The Box above 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.

One aspect of this table that needs to be emphasised concerns partnerships. It is because socio-economic development is so multi-faceted, bringing together different interventions from different policy domains and involving different agencies and actors (formal and informal) that partnerships have become so important in evaluation. Most programmes that derive from the European level are partnership-based as are most nationally and regionally supported programmes. Indeed in Cohesion Policy the Partnership Principle is built into the programme regulations. Partnerships pose particular challenges for evaluation.

Partners always share some common commitments; otherwise they would not become involved in socio-economic development programmes. However, they also have their own interests. These common and specific interests require evaluations and programme managers to incorporate what can sometimes appear to be contradictory objectives, evaluation questions and criteria in their plans. To an extent, evaluations can deal with this phenomenon by carefully incorporating the different voices of partnerships. We will see how this can be done in the next chapter. There are also methods that can be deployed to integrate the multiple criteria of different stakeholders (see Chapter 4). However, working in a partnership context also has implications for the role of the evaluator or evaluation team. They must be more than the passive collectors of stakeholders' diverse priorities. Evaluators need to take on a more active role in order to support consensus building both at the beginning when evaluations are designed and throughout the evaluation cycle including when results are disseminated and conclusions discussed. Co-ordinating, negotiating and facilitating consensus are also necessary skills.

2) The EU policy context

As already noted, any socio-economic development programme is located within a wider policy context. This is true for European Cohesion Policy as it is for programmes at Member State level.

Cohesion Policy programmes have their own orientations and particular horizontal priorities. The Policy is embedded within a wider European policy agenda, which increasingly shapes the content of programmes. So, for the 2007-2013 programming period, the European Council agreed that Structural and Cohesion Fund resources should be mobilised towards the Lisbon objectives for growth and jobs in the context of a concern to maximise sustainable development.

It is important to mention the main policy directions that are shaping approaches to socio-economic development in Europe as these have implications for the overall focus of an evaluation and the particular criteria that might be applied to judge programme success.

Key policy shapers of socio-economic development programmes include:

Community cohesion policy that aims to reduce the disparities, usually measured in terms of per capita GDP, between the most developed and the least developed regions. The policy aims to support those actions that are most likely to contribute to the reduction of the economic, social and territorial disparities in the Union. It does so by concentrating resources in areas that lag behind. The majority of such funds are allocated to regions where GDP per head is less than 75% of EU average. The policy does not address these disparities directly, rather it concentrates on interventions affecting the assumed determinants of economic growth physical infrastructure including transport, human resources and the capacity to manage investments and services especially through a viable SME sector, effective public management and through information technologies.

The Lisbon strategy and process following the special meeting of the European Council in March 2000 agreed a new overall strategy with a ten-year goal of making the Union: the most competitive and dynamic knowledge-based economy in the world by 2010, capable of sustainable economic growth, with more and better jobs and greater social cohesion. Whilst the Lisbon process has continued to evolve (adding environmental, sustainable development and entrepreneurship and competitiveness elements at subsequent European Councils), the central core seeks to integrate well-established strands of European policy. In particular, it brings together the employment and competition policies that can be traced back to the 1993 Delors White Paper on Growth, Competitiveness and Employment (COM [93] 700 final) and social solidarity, welfare and employment systems encompassing what is sometimes known as the European Social Model.

This has been reinforced by the European Employment Strategy[6] which overlaps and reinforces the Lisbon process, for example by emphasising such social dimensions as the reconciliation of work and family life and improving employment for those with disadvantages in the labour market, as well as seeking to improve employability and contribute towards full employment. The European Employment Strategy is also linked with the Structural Funds and there is a commitment to dovetail with Member States Employment Action Plans.

It is these kinds of policy currents that influence the nature of socio-economic development in a European context. Even though there may be differences in emphasis between policies that emphasise solidarity and those which emphasise labour market deregulation, there is a widespread consensus that development needs to encompass both the economic and the social dimensions.

Evaluations of socio-economic development need to take account of this very particular evaluation object. However, the above description of recent policy tendencies also highlights some of the evaluation challenges that are posed by socio-economic development as an evaluation object. For example:

- Many of the policies and programmes (e.g., sustainable development, entrepreneurship, etc.) are complex composite objects. Thus sustainable development brings together social and economic resources as well as natural resources and human skills. The information and knowledge society combines organisational, technological and institutional elements. The evaluation of these raises methodological problems of how to define the object as well as how to know whether improvements or targets have been achieved.
- Within European Cohesion Policy and other socio economic development interventions, horizontal priorities are further reinforced by cross-cutting policies, as many socio-economic initiatives require the integration of resources, measures and implementation mechanisms. These multi-measure initiatives take place in overlapping sites (which may be spatial, institutional or sectoral). Evaluating integrated measures and the interaction of the component parts is part of the complexity that characterises socio-economic development programmes.
- Because of the composite and integrated objects being evaluated, there will often be the problem of choosing the most appropriate unit of analysis. This is partly a question of deciding whether a particular intervention or overall programme or the wider policy is what needs to be evaluated. It also raises questions of aggregation and dis-aggregation: whether to break down complex initiatives into their component parts and judge the success of each or whether to attempt to define an aggregated measure which may not be very meaningful in practice. (For example, can there be a single measure of improved sustainable development?). In practice the choice of unit of analysis is likely to be characterised by compromise and

[6] http://ec.europa.eu/employment_social/employment_strategy/index_en.htm

judgements over trade offs between learning more about specific aspects of interventions and providing insights pertinent to the bigger picture.

- Some of the programmes and policies encountered in socio-economic development can be evaluated within different logics and value systems. In overall terms, for example, should one judge economic disparities at a local, regional, national or European level? In relation to horizontal priorities such as equal opportunities, are they to be judged within a labour market frame of reference or in relation to criteria such as family work life balance or the contributions that these policies make to social solidarity?

Many of the big evaluation questions that can be asked of these and similar policies cannot be answered at a programme level. For example, to estimate the extent of convergence across member states and across regions will usually require Europe-wide comparative analyses of data relating to economic growth, productivity and personal incomes. However, programme evaluations and those of particular interventions or themes can contribute to the interpretation of such larger scale policy evaluations. They can also provide a different kind of evidence: about what is really happening on the ground; about the targeting of measures; about the quality as well as the quantity of programme inputs and outputs; about the reactions and satisfactions of stakeholders; and about the way programmes are managed and implemented.

3) 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 contemporary 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 use of new modes of production consisting of the application of information and knowledge in the production of products and services.
- **Human capital.** Recognises that human resources, in particular literacy rates and education, general health and life expectancy, create conditions for productivity that enable social groups 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,

environmental protection and upgrading, and promotion of equal opportunities. Improved economic and social cohesion has become 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.

- **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.

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 such as cohesion and the Lisbon process 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. The interaction between thematic and regional disparity based methods of allocating resources is also of interest to evaluators of socio-economic programmes as thematic criteria may go counter to targeting those regions and beneficiaries which lag behind the most.

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 infrastructure, skills, investment by firms in new equipment.
- 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 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.

Impact evaluation at the level of an intervention can also contribute to answering the big policy questions. For example, the provision of support to companies to invest in new equipments could be evaluated by tracking the 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?".

4) Community added value

Although the GUIDE is not exclusively concerned with the evaluation of European Cohesion Policy interventions, much of the advice is pertinent to these programmes. Often the evaluations of these programmes and their component parts specifically consider the Community value added of the interventions. There is no right or wrong way of identifying and measuring Community value added and its consideration will need to be tailored to the specific interventions and context. There are however, some starting points:

- Firstly, the principles underpinning the intervention should be considered. These include the principles of: partnership; subsidiarity and additionality. Did the planning and implementation process engage the appropriate partners? Were the key decisions taken at the lowest appropriate level? Were the interventions additiional to what would other wise have occurred?
- Secondly, how does (or did) the intervention contribute to the wider EU policy agenda (Cohesion policy, the Lisbon Strategy, the European Employment Strategy, gender mainstreaming, transport and environmental policies, etc.) should be assessed.
- Thirdly, the extent to which there have been impacts on institutions and systems including the transnational exchange and transfer of good practice or policy mainstreming that are a consequence of EU financing of the intervention should be assessed. (In this respect, as discussed in Chapter 3, EU interventions have themselves been a stimulus for the promotion of an evaluation culture).
- Fourthly, the assessment of Community value added should consider the extent of complementarity between the EU interventions and national policy instruments.

5) 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 always 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, we know the indicators that we would like to collect but the 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, whilst at the same time acknowledging the limitations of what it is possible to say with confidence. The danger is that in response to contractual pressures, evaluators attempt to promise more than they are able to deliver or attempt to 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 at some length in Chapter 3 and in the associated Sourcebook 3[7]. 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 (*see Sourcebook 2)*. 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.

[7] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/capacity_building/index_en.htm

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 have been pulled together below:

1. Remember that we evaluate in order to improve programmes 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. 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. Identify your stakeholders, find out what their interests are in an evaluation and involve them!
4. Evaluations must be fully 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.
5. Getting good work from the diverse groups which make up the contemporary evaluation professional community needs bridge building 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.
6. 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.
7. 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 (see the section of the GUIDE on choosing methods and techniques). 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.

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 3 above), the emphasis on socio-economic development gives particular prominence to one important and often neglected group: 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 and benefits 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. It is also why responsive and participatory methods have become such an important part of the evaluator's toolkit.
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 have 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 planning and project cycle, and as such 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 these documents are often vague, that they try to cover every possible angle of the problem in a sometimes generic way (even if this implies spreading thinly the available resources) and that some of the objectives that they identify are mutually contradictory.

This danger is most present when the complexity of the programmes increases, as with the new generation of socio-economic development policies, stressing the territorial approach, the emphasis on sustainability, the need for extended partnerships and for various mainstreaming principles.

According to one traditional 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 theory, 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 very 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

Importantly, many evaluation techniques (such as evaluability assessment, constructing a programme theory and SWOT analysis) can be used, from a very early stage, in order to clarify the starting priorities and the logic that underpins programme interventions.

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 intermediate 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. This is probably the most important contribution of evaluation to the programming phase.

It is important to note a very important and 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 were the objectives of the exercise? Is this the equilibrium between the different goals that we really want? 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 more comprehensive the indicators system, the more coherent the sequence of the intermediate results and the implementation process put in place, 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 aims (or goals) 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 or has achieved a successful outcome. A further condition which could support rigorous impact evaluation is reflection at the programme design stage on how the counterfactual can be tracked, i.e., what would have happened in the absence of the programme.

In practice these conditions are uncommon 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

2) Planning evaluation

This section more precisely discusses the various activities and issues that are needed when planning evaluation. In particular, it considers:

- Scoping and defining the object of evaluation;
- Identifying and involving stakeholders;
- Analysis of the programme theory 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?

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 and 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.

Activities on the ground impact on the number of stakeholders involved in policy making. In particular, the emphasis on the partnership principle is based on the view that the involvement of non-Governmental institutions and civil society actors will improve the quality of socio-economic development, both from the point of view of defining a comprehensive set of objectives and in terms of facilitating the implementation process.

Other factors which have reinforced the trend towards involvement of large and diverse groups of institutions and actors include the influence of vertical and horizontal partnerships, the emergence of multi-level governance and application of subsidiarity, the establishment of cross-cutting policy priorities such as sustainable development or equal opportunities 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.

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 mid term, 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.

The emphasis on the identification of stakeholders has so far been couched in terms of its practical benefits to understand the programme better, to ask better questions and to obtain good quality information. However, there is an additional rationale for identification and involvement of stakeholders. Evaluators along with 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.

Ideally this exercise should take place before defining the details of the evaluation to be performed: by taking into consideration their points of view, it is possible to decide the most relevant questions that should be answered. But when this is not possible, for instance because it has not been possible to identify all the interested parties at an early stage, some sort of involvement is desirable.

The second question that should be asked is: How is it possible to make sure that the stakeholders provide the relevant inputs to the design, management or content of the evaluative exercise?

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.) as part of the evaluation process. Many methods and techniques described in Sourcebook 2[8] can be used for this purpose: individual interviews, surveys, focus groups, etc.
- At the other end of the continuum, the stakeholders can be involved in steering the whole study, including defining priorities, evaluation questions and associated criteria. Often this means involvement in the Steering Committee for the evaluation project, as we will see when we discuss the management of the evaluation process.

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.

[8] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/index_en.htm

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 desired goal in our case socio-economic development can be reached through the resources allocated and the interventions funded. These assumptions would be consistent with each other and would be supported by evidence. 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 programme theory underpinning the object of evaluation. This is mainly to assess the ability of the programme to reach its intended goals, (i.e., development). A clear identification of the reasons why this should be expected 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 the general goals of the programme, the specific intermediate objectives, the activities put in place by the implementers and finally the results and the consequences of these activities. Chapter 4 (and Sourcebook 2) discusses and provides examples of 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 exercise.

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 programme?)
- *Normative* questions which apply evaluation criteria (are the results and impacts satisfactory in relation to targets, goals, etc?)
- *Predictive* questions, which attempt to anticipate what will happen as a result of planned interventions (will the measures 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 an impact, a group of impacts, 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 final evaluation report 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.

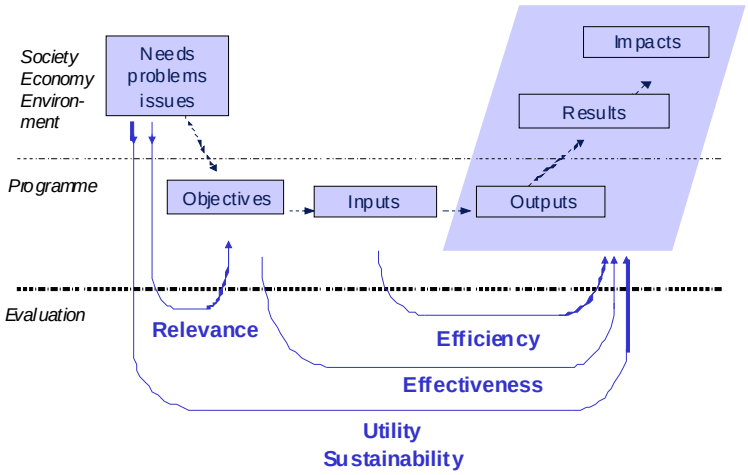
Evaluation criteria

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

- Those related to the relevance of the programme;
- Those related to its effectiveness;
- Those related to its efficiency; and
- Those related to its utility.

These four main categories are represented below.

Selecting priority evaluation questions



The term 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 is supposed to address. In ex ante evaluation, questions of relevance are the most important because the focus is on choosing the best strategy or justifying the one proposed. In intermediate evaluation, the aim is to check whether the socio-economic context has evolved as expected 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 the successes and difficulties have been, and how appropriate the solutions chosen have been and what is the influence of external factors that come from outside the programme.

The term efficiency is assessed by comparing the results obtained or, preferably, the impacts produced, and the resources mobilised. In other words, are the effects obtained commensurate to the inputs? (The terms economy and cost minimisation are sometimes used in much the same way as efficiency).

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

The terms effectiveness and efficiency are commonly used by managers who seek, in the context of monitoring, to make judgements about the outputs (rather than the associated results or impacts). Indeed, questions concerning the performance of a programme are increasingly common within the monitoring framework. Given the relevance to monitoring and evaluation, there is likely to be a fairly large set of questions that will be grouped under the performance heading.

The criterion of *utility* judges the impacts obtained by the programme 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 and outputs 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? Can their raison d'être still be proved? Do they correspond to local, national and European priorities?
- *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 objectives been achieved at the lowest cost? Could better effects be obtained at the same cost?
- *Utility:* Are the expected or unexpected effects globally satisfactory from the point of view of direct or indirect beneficiaries?
- *Sustainability:* Are the results and impacts including institutional changes durable over time? Will the impacts continue if there is no more public funding?

These criteria are not exclusive. Other criteria such as equity, coherence, synergy, reproducibility are also often used in evaluation. In addition, evaluation criteria and evaluation questions that derive from them may relate to the negative and positive unintended consequences of interventions. Even though programmes have their own logics and goals, 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. Nor can evaluation be confined to programme goals and priorities.

Evaluators must also take seriously possible results that were not part of programme architecture. 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 programme effectiveness that may not accord with those of programme architects 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 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 pursuit of learning and programme improvement.

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, results and impacts 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 and results.This may not be the case in smaller socio-economic interventions.

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? This requires that an evaluability study is carried out.

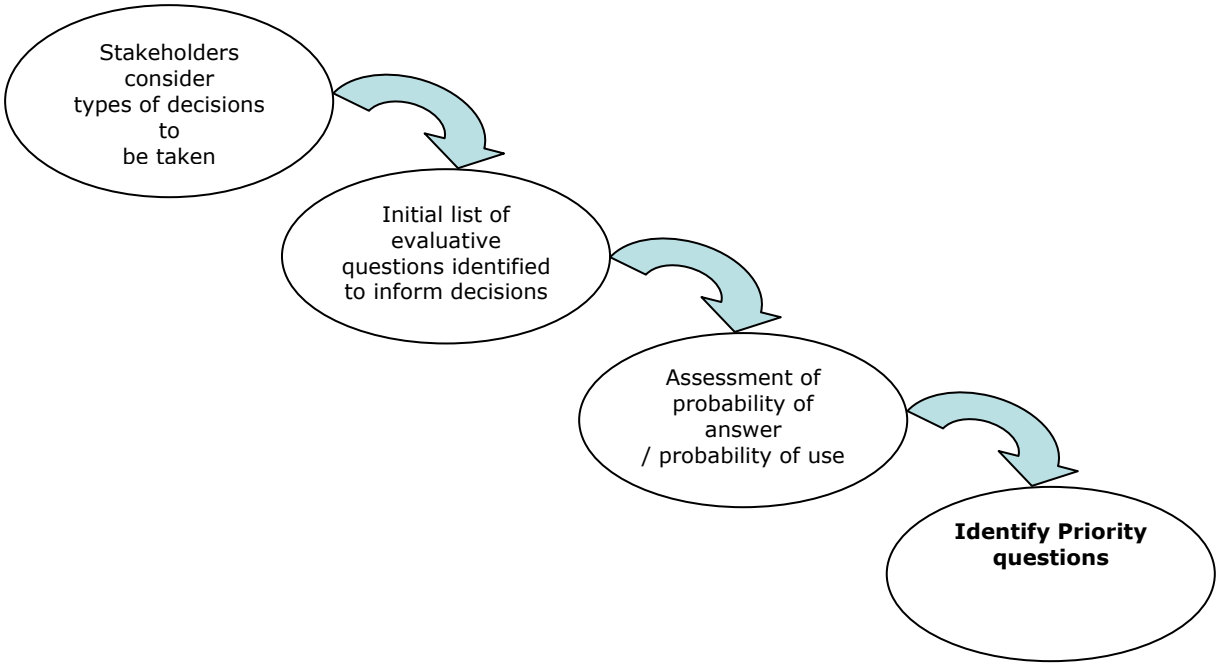
For each evaluative question one needs to check, even very briefly:

- 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. For example, 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 (deciding, debating, informing)? 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?
- Are evaluation priorities sufficiently stable?



In Sourcebook 2 the formal requirements of evaluability assessment are described and those interested in this topic may wish to consult that additional information.

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, it is necessary to make some choices. The choice of methods is 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. Sourcebook 2 elaborates specific methods and techniques, 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 and welcome 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.

Guidelines are especially useful to set the parameters for evaluation in relatively decentralized programmes where it is important that common priorities, criteria and procedures are adopted. This can ensure a degree of comparability. Such guidelines have traditionally been developed by national or European authorities. They can also be useful if developed by programme managers or even evaluators when the overall evaluation process within a single programme is likely to be decentralized. In socio-economic development when participative evaluations and self-evaluations are common, some basic guidance, especially if priorities are developed collaboratively with local actors, can be most effective.

For the 2007-2013 programming period, the obligations for evaluation during the programming period have been 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 will require 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.

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%). There are incidences where up to 5% of programme budgets have been devoted to evaluations that are effectively part of management's implementation strategy. For example, where evaluation includes a strong formative element intended to assist managers and stakeholders with their work.

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.

In practice it is common in socio-economic programmes to spend sums unnecessarily when evaluations address routine topics but not to spend enough when programmes are innovative. This is, of course, the danger when evaluation is intended primarily for accountability or monitoring purposes.

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 mid term 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 on going 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) can be 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 of impacts and results.

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 (descriptive, causal, critical, prescriptive or normative) 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 imagine 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.

- Independence of the evaluation team. 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. (Issues of evaluator independence are discussed in greater detail below).

- Profile of the evaluation team

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 significant 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. On the other hand, 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.

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?		++	

Above all it must be remembered that judgements on the quality of the method proposed are qualitative and not quantitative. These are judgements, which need to be made by those with experience. Many of the items for which judgement has to be made are also qualitative. 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.

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 or do not completely fulfill the specifications in the ToR. It could be argued that if the evaluation is standard or the ToR are very precise about what must be done, the quality of the personnel and the price are the only things that actually matter. However, although this may be so, the reverse is not true. When the tender process asks candidates to propose the methodology that they consider suitable for the task the danger is that too much attention is paid to the originality of the approach and not enough to the ability of the candidates to actually deliver what they have promised.

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. As a rule of thumb for any evaluations that are not entirely routine, the financial criterion should not exceed 20-25% of the overall assessment.

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 better 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 mere 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 ongoing 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. Even in self evaluation approaches those involved in the implementation of interventions need to achieve a degree of distance and independence whether or not they are assisted in the process by outside evaluators. 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 whole 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 well 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 substantive 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 implementation of evaluations. 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 *a priori* 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 *a priori* 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 true 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, in particular in all interim, ex post or thematic evaluation exercises.

An inception or start-up phase will usually aim to 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 operational approach 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. The Box below provides an example.

Czech Republic

In the context of the ex-ante evaluation of the National Development Plan in the Czech Republic for the 2004-2006 period it was noted that the recommendations of the evaluation team proved to be more readily accepted if they were suggested in the early stages of drafting programming documents. The work of the evaluators had developmental features often associated with technical assistance and implies that the evaluation was perceived as more useful when real partnership had been established.

Even if the evaluation exercise is straightforward, external policy contexts often change rapidly. It is therefore useful to secure effective liaison not only with the members of the Steering Committee but also with those responsible for policy-making. The opportunity to involve, whenever possible and even indirectly, the strategic level of management is another reason why the process must be interactive.

One simple mechanism is to specify the frequency of Steering Committee meetings even at the ToR stage. A minimum of two meetings are usual at inception and to approve a draft final report.

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 several weeks in order to plan and assemble the team that will actually do the work. There are at least two ways in which this flexibility can be guaranteed:

- delay the actual signature of the contract, and therefore the starting date of the job;
- allow an adequate period for the Inception Report.

Role of Inception Report

The Inception Report is a document which 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 an 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. In Cohesion Policy, this interim stage is often included in interim or ex post evaluations. Ongoing evaluations that track changes over time typically have a series of interim reports that provide feedback to programme managers and policy makers.

Draft final reports can perform a similar steering function, especially if they are required early enough. However, these mainly steer the report, rather than the programme, which would be the case with interim and ongoing evaluations. It needs to be emphasised that in the interests of independence, Steering Committees that receive draft final reports 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. A multi-level approach to involving strategic management on the Steering Committee is important as programmes grow increasingly complex when they take into account concerns that have different territorial dimensions;
- *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. Even so, it is an important task of Committee chairpersons to ensure that no members, including operational managers, 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 principal 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 should 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 major 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 (it is not unusual for interim reports to 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. Interim findings can provide useful feedback on process and implementation (e.g. suggest changes in procedure) and help increase the engagement of stakeholders and those involved both in the programmes and in the evaluation.

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 an abstract anticipation of outputs 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, increasingly 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. Verbal 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. Indeed an evaluation that does not meet some minimum quality standards can very well mislead decision-makers and programme managers.

However, to assess evaluation quality is a complex and difficult 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 monitoring system and 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.

Furthermore these aspects can be seen from two different points of view.

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 and the reverse is also true, in the sense that a good product should be the result of a good enough production process.

The MEANS Collection (1999) noted:

There is no system of professional certification, anywhere in the world, which institutionalises rules and quality criteria. Somewhat disparate grids of criteria are proposed, based on the evaluation models elaborated by various authors, but no consensus exists in this domain. Moreover, the nature of the criteria mentioned does not always appear clearly when an attempt is made to put them into practice.

Since then, however, some things have improved. 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.wmich.edu/evalctr/jc/PGMSTNDS-SUM.htm

- Deutsche Gesellschaft fuer Evaluation (DeGEval): Standards fuer Evaluation (2001) : http://www.degeval.de/index.php?class=Calimero_Webpage&id=9023&PHPSESSID=10cc3ea5a97acc7654ac37a5ed846520

- Société canadienne d'évaluation (SC) Lignes directrices en matière d'éthique / Guidelines for Ethical Conduct : http://www.evaluationcanada.ca/site.cgi?s=1

- Switzerland: SEVAL Evaluation Standards: http://www.seval.ch/de/documents/seval_Standards_2001_dt.pdf

- The African Evaluation Guidelines 2000: http://www.geocities.com/afreval/documents/aeg.htm

- 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/Pub_library/PRO2907%20-%20UKE.%20A5%20Guideline.pdf

- 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

The list above provides a cross section of some current evaluation standards and codes. They fall into a number of categories. 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, e.g., they concern how data is gathered and how conclusions are presented. (The distinction between guidelines and the more stringent and ambitious standards is also instructive.) Another category, of which Canadian and Australasian and to some extent the UK Evaluation Society's outputs are examples, is more concerned with ethical codes of practice than technical practice of evaluation. But again this mainly concerns the ethics of evaluators than of other implicated actors. Most recently a new category of guideline has emerged. This is directed more at administrations and those who commission evaluations than at evaluators. Examples of this can be found in the OECD (PUMA and DAC guidelines) and most recently in the European Commission.

Despite this growing array of guidelines, standards and codes that concern quality in evaluation there is not at present a common statement that has universal recognition.

Although there is not yet consensus about all the components of a quality assurance system for evaluation, we have begun to see a shift from a focus largely on quality control, i.e., ways of judging report/ output quality. This shift was endorsed by a recent study on the use of evaluation by the European Commission (see Box below).

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).

The Quality control and quality assurance criteria identified below include both quality control and quality assurance criteria. Both are needed as a means of judging evaluation reports and outputs. Normally the person responsible for managing the evaluation within the commissioning body would take responsibility for applying the quality control criteria. Ideally performance on the quality assurance criteria needs to be informed by the views of members of the Steering Committee, other stakeholders, the evaluation team and those responsible for managing the evaluation on behalf of 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. However, for quality assurance that rests on process criteria, consultation with other stakeholders not necessarily represented on a steering committee will be necessary. For quality control purposes, consultation with external experts or referees can be useful. It needs to be emphasised that the application of quality control / content-type criteria and quality assurance / process-type criteria are undertaken for different purposes. Quality control of report content offers some 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

Who should be responsible for a quality control and quality assurance procedure will vary with the institutional context. In national sectoral programmes, this may be a central government responsibility and in local development programmes, the responsibility may rest with local actors. The methods of application will be similarly varied - sometimes a grid may be filled out by key individuals and aggregated, but on other occasions a workshop or consensus conference may ensure the most balanced judgements.

Quality control - output criteria

- Meeting needs

Has the evaluation answered the questions included in 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 interpreted 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?

For ex post evaluations it is important to check whether the evaluation has managed to reach a reasonable compromise between the following two contradictory requirements: rapidly obtaining information for feeding into the new programme cycle and not drawing hasty conclusions before all the impacts have been observed.

▪ Relevant scope

In order to check the relevance of the scope of an evaluation, it is necessary first to check whether the essential characteristics of the programme or intervention have been well described and whether the problems and successes in implementation have been properly clarified.

Secondly, because the results and impacts have to be analysed in order to judge the extent to which objectives have been achieved, it is necessary to check whether they have been included in the evaluation. It is also necessary to check whether the evaluation has overlooked other potential or future results or impacts, as well as any unexpected yet significant effects and results that may exist.

Finally, the scope of an evaluation depends on the programme or intervention target that can be defined in terms of eligible geographical areas or non-localised target groups (e.g., the long-term unemployed). It is therefore 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;
- lastly, if the evaluation considers the evaluated programme or intervention in isolation or includes its interactions with other European or national programmes.

▪ Defensible design

This criterion relates to the technical qualities of the evaluation. Methodological choices must be derived from the evaluative questions. The evaluation must, moreover, make the best possible use of existing research and analyses. Three types of question have to be asked:

- Has the relevant knowledge been collected and used wisely?
- Are the construction of the method and the choice of tools really justified for answering the evaluative questions properly?
- Were the reference situations chosen (counterfactual or similar) appropriate for making valid comparisons?

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

- the method is described in enough detail for the quality to be judged;

- the validity of data collected and tools used is clearly indicated;
- the available data correspond to the tools used.

Because a causal analysis of effects is the most important question in ex post evaluations, the method used to analyse these causal relations is the priority in this type of evaluation. It is necessary to check whether the evaluation adequately analyses relations of cause and effect for the most essential questions.

- Reliable data

Evaluators use existing data (secondary data) from the monitoring system and from other sources of information, or else primary data that they have collected for the evaluation. In the latter case, the methods used to collect and process the data (choice and application of the tools used for this purpose) are very important factors in the reliability and validity of the results.

In order to assess the reliability of the data used, it is necessary to examine whether:

- available sources of information have been identified and the reliability of this data has been checked;
- sources of information taken from the monitoring system and previous studies have been used optimally;
- the techniques used to collect the chosen data were complete and suitable for answering the evaluative questions.

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

- the mixture of qualitative and quantitative data is appropriate for a valid analysis of the phenomenon;
- the "populations" used for data collection have been correctly defined;
- the survey samples or cases studied have been selected in relation to established criteria;
- the main data collection techniques have been implemented with appropriate tools and in such a way as to guarantee an adequate degree of reliability and validity of the results.

- 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 in the form of cross-referencing with a particular focus on why things happen. In both cases it is necessary to assess whether the methods of analysis used are relevant as regards the type of data collected and whether the analysis has been carried out to an appropriate quality.

In the case of socio economic development 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 sufficiently 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.

▪ Credible results

The credibility of results is defined here as that they follow logically and are justified by the analysis of data and interpretations based on carefully presented explanatory hypotheses. The validity of the results must be satisfactory. This 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. It is also necessary to check whether the results of the analysis were produced in a balanced and reliable way.

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 that are more than results. Whereas results are "technical" and can be analysed without too much risk of impartiality, conclusions and, *a fortiori*, 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.

Key questions such as relevance, effectiveness and efficiency must be addressed within the framework of an evaluation and must therefore be answered appropriately. The evaluation report must also show the appropriateness of the budget for the programme of intervention.

Essential questions such as the value added of the programme or intervention and progress made in terms of transversal goals like cohesion, subsidiarity, good governance, sustainable development and equal opportunities may need to be studied. In the case of ex ante evaluations of programmes, conclusions need to be formulated so as to feed into the process of negotiation on the evaluated programme. The report should make it possible to improve the evaluability of the programme or intervention.

- 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 absolutely 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. It is therefore essential for this summary to 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 outputs 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. To an extent it is possible at the time they are drafted to judge the adequacy of a ToR. It also becomes easier with hindsight to identify what might have usefully been included. This is important for future learning, i.e., how to improve ToRs in the future.

A poor or incomplete ToR can lead evaluators to deploy their resources inappropriately. It can also lead to other negative effects. One common consequence is when gaps in the ToR become evident in the course of an evaluation and the commissioner struggles to redirect the evaluation mid-way or to request additional outputs that were not planned for or budgeted.

- 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. For example, was a decision taken to favour a well-known firm but the time commitment of key personnel was inadequate? Was the method too loosely specified? Or was an experimental high-risk method favoured and could this account for problems encountered later?

- 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 in evaluation. This is partly a question of the forum created for this purpose. Most obviously a Steering Committee but possibly also specific briefing meetings and workshops, e.g., briefing workshops for local politicians and policy makers. The inclusiveness of the membership of such meeting places needs to be assessed: were all the right stakeholders and publics involved?

The purpose of these opportunities for briefing and exchange is the dialogue and feedback that they enable. 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. Chapter 4 emphasises the importance of data availability and monitoring systems. Without adequate information resources it is difficult for evaluators to do good work. An assessment therefore needs to be made of the adequacy of information. Most obviously this concerns monitoring information and systems. Often monitoring systems emphasise the needs of external sponsors and funders. They also need to be able to 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 programme but also how to interpret more formal data.

Overall, in order 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. For example, in some programmes there may be data available such as administrative returns on local employment or the minutes of management committees of particular projects or sub-programmes but these are difficult to access. It may also be that the key informant refuses to provide evaluators with information perhaps because of poor relations between the involved stakeholders and administrations. To that extent, 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 their outputs integrated. Relations with commissioners of evaluation, programme managers and a whole variety of informants, fieldsites, 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 there are also elements that are shared with programme managers and those who are responsible for commissioning the evaluation. For example, how the commissioning system responds to requests to adapt a previously made work plan is not in the control of the evaluation team alone.

- 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. (This is covered in terms of quality control above.) It also requires an awareness of the value and opportunities for dissemination within the evaluation team. There is for example a big difference between evaluators who confine their feedback to the contractual minimum and those who see it as their responsibility to provide ad hoc feedback when new problems occur or when key issues need to be resolved.

This kind of dissemination also requires sensitivity to the information needs and interests of key stakeholders. Sometimes outputs need to be tailored to meet quite different interests. For example programme managers will have a different perspective from local SMEs even though they will also share certain interests in common.

- Effective dissemination to stakeholders

Reports and outputs need to be disseminated if they are to facilitate learning. 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. However, primary responsibility does not rest with evaluators. Programme managers and those who commission evaluations should take responsibility for dissemination to stakeholders, including the public at large.

▪ The synthetic assessment

The synthetic assessment recapitulates all the above quality criteria. It is difficult to recommend any particular weighting for the different criteria because their importance varies from one situation to the next.

Below is a synthetic grid for an assessment of the quality of the evaluation:

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:	☐	☐	☐	☐	☐

A five point rating scale is used. This runs from the positive (where very positive indicates the end point) to the negative (where very negative indicates the end point). Thus there are two positive possibilities and two negative possibilities and a mid-point when the balance of judgement is uncertain.

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 or 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, clarify thinking and develop frameworks.

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 and directors. 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 application of the process is uncommon.
- 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. (Institutional arrangements are discussed further in Chapter 3).

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 programme planners. In particular, this can help to clarify objectives of programmes and interventions. 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. Various techniques such as evaluability assessment and preparing an analysis of programme theory can be deployed for this purpose. In general, in order to ensure independence of the evaluation proper, it would be best to use different evaluation teams or resources for this programme planning work than for the main evaluation.
2. A similar evaluability assessment should be undertaken by evaluators when they begin their work. To some extent this may overlap or repeat what has already taken place at planning stage. However, the purpose here 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 and agendas. It will also ensure that they feel some sense of ownership of the outputs of the evaluation and are more likely, therefore, to find these useful and use them. On the other hand, it may be necessary to be selective in deciding on which voices finally determine the evaluation agenda, in order 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 that need to be made along the way. It will also ensure 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 will be 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 and helping those for whom it is intended? Does it support equity or not? Is the programme consistent with other policy initiatives? And is it delivered in an efficient and legitimate way? Maintaining this broader perspective ensures that for part of their work at least, evaluators are able to 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. You should try to 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 is important therefore not to simply follow a standard framework with pre-drafted paragraphs. Rather 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 that are 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. So-called process or dialogue use should involve stakeholders in evaluation thinking from the beginning. There are evaluations where the conclusions and recommendations are rejected but stakeholders, especially the core stakeholders involved in the steering committee, nonetheless find the evaluation useful. It can help them to clarify their own thinking and understanding and spark off innovative ideas for improvements. This continuous process of communication provides a particular context for the dissemination of evaluation reports and findings. 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 of the kind that has been advocated in this Chapter provides a context in which to explain the strengths and weaknesses of evaluation outputs. It also offers an opportunity for those who commission evaluations to learn how to improve evaluations in future.

10. Consideration should be given at an early stage to how evaluation findings will be put to use. 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

Introduction

This Chapter discusses one of the pre-conditions for the evaluation of socio-economic development: having and developing the capacity to undertake evaluations. It begins by discussing the importance of evaluation capacity and what it is. This is followed by an outline of the actions that should be considered in order to strengthen capacity including various strategies that have been found to be effective in different national settings. Evaluation capacity is discussed from both a demand and supply perspective: the need for administrations to request, commission and use evaluations and the need for people, skills and institutions to be available that can respond to these demands. The starting point is the needs of national public management but the role of decentralised administrations and local programme managers is also recognized as well as civil society including consultants, universities, communities and the private sector. The contents of this Chapter are supported by one of the associated Sourcebooks, which includes case studies of the way in which various countries have over the years strengthened their evaluation capacity.

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 associated 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 without costs. However, the potential benefits of evaluation are large enough to justify the initial 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 cannot be put in place at a single moment in time. They 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. As has been emphasised many times in this GUIDE, independence is an essential ingredient of evaluation. To begin with this 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 is not a stand-alone function: it requires an institutional framework and has to be embedded in institutional arrangements. In particular 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. Indeed, 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 the 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 and separateness 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:

- First, the systematic identification of the best alternatives, as well as the careful consideration of the ability of ongoing, past and indeed future programmes to reach their objectives in an efficient way. This was identified in the previous Chapter as an important contribution of evaluation, which can become a powerful tool for modernisation in the public sector, for cost reduction and for greater responsiveness to citizens.
- Second, the opening up of the administrative black box to the scrutiny of external stakeholders, as well as taking the interests of stakeholders and citizens into account (e.g., by including civil society groups in steering committees, consultations on evaluation findings and conclusions, etc.) is in itself an embodiment of the principles of democratic governance. Because the pledge to systematically evaluate public programmes can enhance the trust of citizens towards government, it contributes to the increase and maintenance of social capital. As contemporary theories of socio-economic development rest heavily on mobilising territorially-based (endogenous) resources and potential, such increases in social capital help to promote sustainable socio-economic development.

What is evaluation capacity?

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

- Individual level consisting of necessary skills and competencies;
- Organizational level of management arrangements and structures;
- Inter-organizational level that bridges public and private bodies through networks, procedures and partnerships; and
- 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.

There are many indicators of evaluation capacity, mostly arrangements that can be expected to be put in place for an evaluation system to be sustainable. Whilst it is not feasible to try and introduce all of these arrangements at the same time, an ideal end-state is worth describing so as to allow for a planned strategy for implementation. In such an ideal situation one would expect the many of the following arrangements to be found even though the following examples are not exhaustive or comprehensive:

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 and training and professional development.
- Training courses and diplomas are available nationally or on a regional basis, variously delivered by universities, private training providers and professional bodies and open to both 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 deliver and when identifying lessons about 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 sectoral and technical expertise and understand the priorities of policy and programme customers.

- 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 and consultants 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 and plan if evaluation capacity is to become useful within a defined time-scale. This has been the case in most new member states of the European Union. 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.

Evaluation capacity in Lithuania:

Lithuania had little prior experience of evaluation prior to accession and limited capacity for the evaluation of structural funds. The chosen strategy was to locate overall responsibility for evaluation with the Ministry of Finance, whilst involving all stakeholders in an Evaluation Management Group. A consortium of foreign and local consultants was contracted to undertake an evaluation capacity building project.
Among the measures that followed were:

- **A regional conference on evaluation organised to raise awareness**
- **Evaluation training for civil servants and experts**
- **Guidance and methodological documents were prepared**

Source: Conference Presentation, Director EU Programme Management Department Ministry of Finance, Lithuania

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 and protocols, personnel and 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 and concentrates its work there. 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. See example in Box below:

Coordination of Evaluation in Romania:

The Managing Authority Evaluation Unit in Romania has a number of key tasks, including to:

- **Chair and act as Secretariat to a cross Managing Authority Evaluation Working Group**
- **Coordinate and manage the evaluation within the national framework (NRSF)**
- **Oversee the drafting of a national evaluation (capacity) strategy**

Source: Interviews with Romanian authorities

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. A recent World Bank publication suggests that this has happened in some countries: centralisation led to little use of evaluation elsewhere in the government concerned. Too much de-centralisation on the other hand - for example having different coordination arrangements for different programmes, policies or territories, risks incoherence.

Confining the early stages of evaluation capacity development to a pilot ministry risks the familiar problem of innovation enclaves. The location of evaluation development might adapt too closely to a single policy environment and prove inflexible and difficult to transfer. This has been an issue in some member states which have invested heavily in evaluation capacity development in Cohesion Policy and have then found it difficult to extend evaluation systems and practices to other nationally

financed programmes. There have even been cases of accession States that have adapted their evaluation systems to PHARE (pre-accession and transition programmes) requirements and have found it difficult to transfer models, skills and institutional arrangements to Cohesion Policy. As the Irish case indicates, it is not uncommon for evaluation units to be restricted to EU funded programmes at the beginning and only later have their mandate extended to programmes funded through national Treasuries.

Experience of Interim evaluations in Ireland:

In Ireland between 1992 and 1995 the ESF Programme Evaluation Unit, the Industry Evaluation Unit and the Agricultural Evaluation Unit were created with independent status. Most of the managers and some of the evaluators of the units came from government departments or associated agencies and therefore knew the system well. This closeness facilitated access to information and in general the units enjoyed free access to information in departments. In terms of internal versus external evaluation the units may be said to be a hybrid that attempted to capture elements of both. The units were mainly responsible for carrying out interim evaluation where the main focus was on managerial aspects of the programme.

Source: The Irish case-study in Source Book 3 Capacity Building

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. An example of this would be a network of evaluation advisors - one in each Cohesion Policy Operational Programme but coordinated across all OPs by a central Unit.

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 the European Union and in particular 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 as it is for most of the new Member States. This external demand is in part an outsourcing of the responsibilities of European institutions that are obliged to demonstrate that they are 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 and social 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, good practice and value for money, i.e., they are interested in learning, as well as accountability. 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 (see Box below).

This does not mean that even domestically evaluation is not sometimes externalised. 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 include notions of effectiveness, efficiency and value for money;
- A finance minister wanting reports from the departments to which budgetary allocations have been made;
- 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 self interested 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.

Evaluation culture in the Netherlands:

A lesson that it is possible to draw from the experience in the Netherlands is that, for a culture of evaluation to develop within a country it is important that the motivation for carrying out and using evaluation is not merely externally driven. The internal motivation of Dutch government to improve public policy, motivated by notions of credibility and cost effectiveness of public management can indeed be seen as an important factor for the relative success in introducing and using the result based management model.
Source: The Netherlands case-study in *Sourcebook 3 Capacity Building*

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.
- As a correlate of the 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 first, 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. For example, many of the specific activities identified in Chapter 2, concerned with the commissioning and management of evaluation become relevant. Establishing procedures for consultation with policy makers and managers when devising an annual evaluation plan or terms of reference; setting up steering committees; reviewing the quality of evaluation reports; all these processes can generate secondary demand for evaluation. However the strongest boost will come from embedding evaluation in the routine practices of public agencies or government ministries. This is discussed in greater detail.

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 themselves engaged with public policy making and service delivery and have a responsibility for efficiency and responsiveness they share with governments. 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.

An interesting example of this within the European Union is in the Basque region of Spain, where a deliberate attempt has been made in relation to developing evaluation of evaluations of the Basque sustainable rural development plan (within EU Structural Funds) to adopt a participative evaluation design.

Participative evaluation capacity:

The evaluation should, in the first place, be a useful practice for those in charge of running the programme; however, it should also provide the same function for all the potential users and collectives affected by its outcomes; beneficiaries, social bodies, civil associations, etc. From this perspective, evaluation thus becomes an exercise aimed at responding to the need for information from the social agents as a whole and even goes as far as to promote their active participation in the evaluation process.

Source: Diez, Malagon and Izquierdo

Quoting a classification of evaluation capacities undertaken by C3E in Lyon, the authors suggest that this type of evaluation is typical of a developed stage of evaluation.

Civil society and evaluation demand:

[Evaluation] has begun to be a social and political act and its conclusions are the subject of public debate (political use of evaluation). The objective is to provide the public with information about the efficacy of the use of public resources by the administration. In this context evaluation becomes an instrument in the democratic process in that it informs citizens, delivers accountability and gains social approval for state interventions. Countries in which the development of evaluations has reached this stage are countries such as Denmark, Holland, Sweden and the United Kingdom.

Source: Diez, Malagon and Izquierdo based on a typology developed by C3E Lyon as part of the MEANS Programme

It is certainly the case that in the countries referred to, evaluation is conducted first in response to internal demands much more than to meet external (e.g., EU) accountability requirements and these evaluation procedures often pre-dated or at least developed independently from Cohesion Policy regulations and associated evaluation systems. However even more striking is the decentralised, multi-level nature of evaluation, mirroring the realities of multi-level governance.

In the UK, for example, it is common to complement national, policy evaluations with local evaluations that are intended not only to consider decentralised issues of delivery and implementation but also to answer evaluation questions that are of interest to policy and programme beneficiaries and local communities and NGOs as well as national policy actors (Biott and Cook, 2000). Civil society actors once mobilised also become a source of evaluation demand if they can see potential benefits for themselves.

Supply side

1) Starting Conditions and Traditions

However strong the demand for evaluation, the response - and speed of response - to that demand will depend on pre-existing resources, skills and institutions. In part as already emphasised this will be a matter of intermediate supply embodied in government and administrative routines. But to a

greater 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, training and university 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 establishments and institutes already have much of the underpinning 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 one or more centres of excellence linked to research (e.g., into public management or regional development) and able to disseminate knowledge. Courses can be directed at different target groups. For example, 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.. Specific issues, relevant to the field of socio-economic development (e.g., transversal priorities such as information society, equal opportunities), should be included within these programmes

It is important when planning education and training programmes to recognise practitioner aspects of evaluation in the socio-economic domain. 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 an academic environment. Practitioner skills are needed as well as academic knowledge. This can be achieved by including within curricula that involve less formal teaching and learning approaches, for example, workshops, training seminars, guest lectures from practitioners, work placements with existing evaluation consultants, study visits to administrative bodies with responsibility for socio-economic development, etc..

In some countries there is a tradition of universities and independent institutes actively engaging in policy and practice related 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 workshops and 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. This is so for socio-economic development evaluations. 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 usually thought 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. This practice was also encouraged in the EU with the broad notion of twinning between EU15 and new member states and more specifically reflected in the composition of evaluation teams made up of similar international partnerships especially when contracting for multinational, regional and sectoral evaluations in the years preceding accession.

Some new consultants will come from or be partially based in universities but others will not. It is also a characteristic of new approaches to public management that this will be associated with the creation of a number of new markets for consultancy services. To that extent promoting the independent consultancy sector will have wider benefits than for evaluation alone.

Evaluation capacity development in the Czech Republic:

In the Czech Republic, an expert base existed in particular in the academic community that could be used for the purposes of evaluation. However, there were no Czech companies specializing in the supply of evaluations, the offers of foreign companies were too expensive, and knowledge of the Czech environment was inadequate. In the practical sphere, many domestic entities had experience in preparing plans, programmes and in drafting policies to support socio-economic development. These included the National Development Plan and the operational programmes. This practical experience from a related area of evaluation is very useful for the launch of evaluation capacity building.

Given the high cost of evaluations, the need for knowledge of the Czech environment and the requirements related to the development of evaluation capacity in the Czech Republic, in contracting ex ante evaluations to external evaluators preference was given to domestic contractors. These contractors were new in the field compared with their foreign counterparts, whose services could not be used mainly because they were too costly and also because there did not always have a sufficiently rounded understanding of the Czech environment. In several cases, the Czech companies commissioned to conduct the evaluation took on a foreign expert to help pass on unwritten, but generally applied evaluation processes and to ensure that the European standard was respected. These consortiums fulfilled what was expected of them in their practical activities. A positive finding was that the operations of purely Czech companies, which did not incorporate any foreign experts, produced much the same results.

From: Czech Community Support Framework Evaluation: Mario Hladek, David Grolig (with Eva Pisova Head of Evaluation Unit) Prague, September 2004

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;
- Support for networking among relatively isolated evaluation consultants so as to encourage team-building, consortia formation and other professional networks and associations (see below); and
- Acknowledgement by evaluation commissioners that they may need to take a more hands-on management of new contractors to speed up their acquisition of the tacit knowledge that experienced consultants already have.

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 commonly 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.

Evaluation Standards:

[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. DG Budget October 2002

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. (Material is available on national and regional evaluation societies on several international evaluation websites. See especially IOCE and IDEAS.) These typically bring together people from different institutions (academia, public administration, research centres and consultancies) with different disciplinary expertise (sociologists, economists, planners, political scientists, pedagogues, 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 (see below for a list of some of the current European evaluation societies). 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. As was noted in Chapter 1, most of the evaluation standards and codes of ethics that have emerged in recent years have come from the work of evaluation societies and associations.

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.finnishevaluationsociety.net/**
- **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.idr.es/idr_in/index_intro.html**
- **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.

Defining institutionalisation:

Institutionalisation in evaluation is often understood as both rule-based and as a process. According to Boyle, Lemaire and Rist, it is the rules, procedures and organisational arrangements by which the evaluations of public policy are produced and the process of legitimization by which evaluation practice formally becomes part of the decision making process of government

Building Effective Evaluation Capacity, Boyle and Lemaire (Eds) Transaction Publishers 1999

It is likely that the introduction of evaluation will first be concentrated on one part of a government, a single ministry or a set of programmes. 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 and prioritisation within public agencies 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 or are not following recommendations;
- Positive incentives for managers to use evaluation findings and modify programmes and their direction if necessary;
- 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.

Planning for evaluation extension and deepening in Romania:

In Romania leading up to accession evaluation efforts were initially concentrated on the financing, planning and design stages of new European policies and programmes. Efforts also first focused on PHARE programmes and preparation for Structural Funds. However plans were made to develop an evaluation culture that would be relevant to all all EU funded Structural Fund programmes. The central coordinating Ministry was in parallel orchestrating a modernisation and reform programme across all ministries. This was also linked with plans for eventually extending evaluation to nationally funded programmes and policies

Source: interviews with Romanian officials

Creating an evaluation culture is a phrase that is intended to suggest 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 also 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.

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

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[9]:

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 key component in public management this differs for example to a purely backward looking accountability focus, although accountability is also emphasised

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

Excerpts from Evaluation Policy: Treasury Board of Canada:

***The policy is based on three fundamental principles*:**
- that achieving and accurately reporting on results is a primary responsibility of public service managers;
- that rigorous and objective evaluation is an important tool in helping managers to manage for results; and
- that departments with the support of the Treasury Board Secretariat, are responsible to ensure that the rigour and discipline of evaluation are sufficiently deployed within their jurisdictions.

Managing for Results
Management for results is the prime responsibility of public service managers.. managers must be accountable for their performance to higher management, to ministers, to parliament and to Canadians

Evaluation as a Management Tool
Evaluation operates in a complex environment that involves partnerships with other federal organisations, with other levels of government, with the private sector or with not-for-profit entities. In addition, as the delivery of programs is devolved to other jurisdictions, the evaluation of underlying policies increases in importance. Evaluation should contribute to improvements in policy, as well as program design and delivery.

In this milieu, departments should embed the discipline of evaluation into the lifecycle management of policies, programs and initiatives to:

- **develop results-based management and accountability frameworks for new or renewed policies, programs and initiatives;**
- **establish ongoing performance monitoring and performance measurement practices;**
- **evaluate issues related to the early implementation and administration of the policy, program or initiative, including those that are delivered through partnership arrangements (formative or mid-term evaluation); and**
- **evaluate issues related to relevance, results and cost-effectiveness.**

Various diagnostic tools have been developed to assess the degree of institutionalisation of evaluation and provide guidance for how this might be strengthened. In Lithuania for example OECD typologies were adapted to assess the degree of evaluation institutionalisation. Factors considered by Nakrosis and colleagues from the Public Policy and Management Institute included:

- Evaluation timing application of evaluation at all policy stages,
- Degree of centralisation,
- Location of evaluators inside or external,
- Integration with decision making.

Such applications of standards allows for the identification of strengths and weaknesses and a plan for improvement.

The World Bank has produced a framework for assessing options under different conditions. It starts from the recognition that in any setting of national administration the starting conditions for evaluation capacity development will vary. In some demand, especially internally driven demand,

will be stronger and in others less so. In some supply, the availability of skills and networks of institutions able to mobilise these skills, will be more capable of responding to evaluation demand than in others. Whatever the balance of supply and demand, there are appropriate strategies that can be adopted to reflect prevalent circumstances.

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:

- First, as this Chapter has emphasised, 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.
- Second, 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.
- Third, 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, the 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 Sourcebook 2[10]. 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 broad families of methods and about specific techniques within these families. 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. This Chapter begins by considering: the context within which choices over methods and techniques are made; the somewhat blurred distinction between quantitative and qualitative data; and the ways in which evidence is obtained and used. The subsequent sections consider:

- The methods and techniques applicable for different types of socio economic interventions, including the main thematic priorities and different policy areas.
- The methods and techniques for different evaluation purposes (planning, accountability, implementation, knowledge production, institutional strengthening)
- The methods and techniques applicable to different programme and policy stages, from policy formulation through to impacts
- The methods and techniques applied at different stages in the evaluation process (Sourcebook 2 presents methods and techniques in this way)
- Acquiring and using data with different characteristics in evaluation
- Creating indicators and indicator systems
- Using indicator to improve management
- Golden Rules

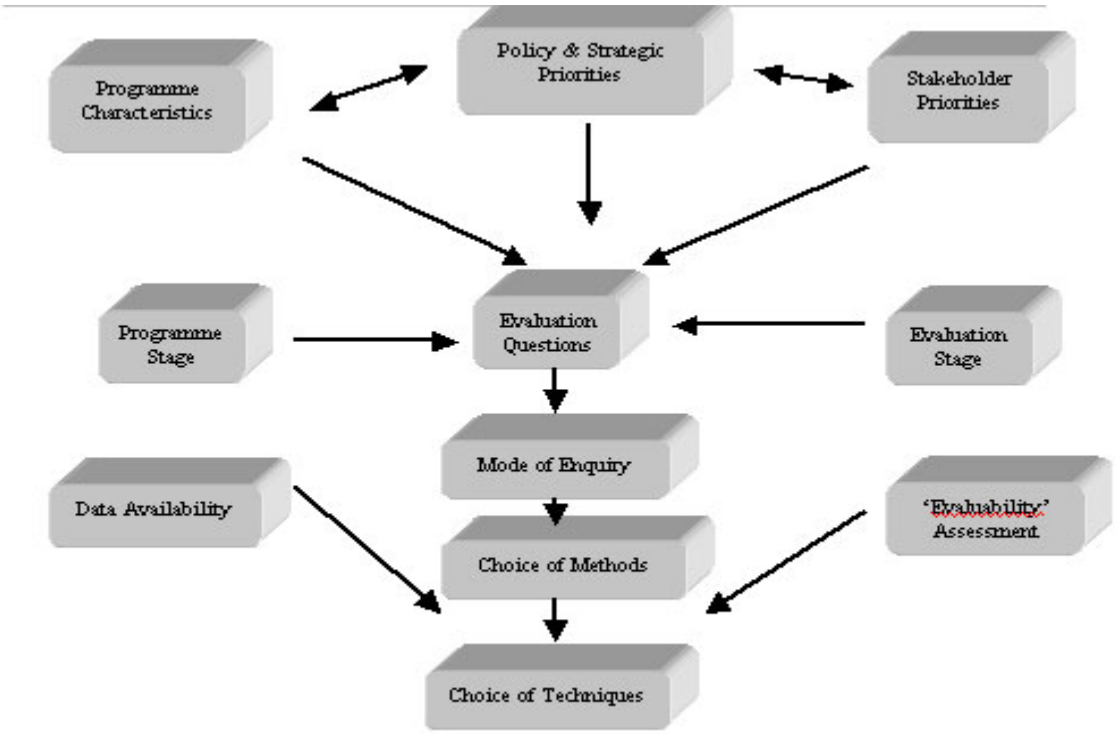
[10] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/index_en.htm

Factors influencing the choice

1) Choosing methods and techniques

As been discussed in Chapters 2 and 3, there are many decisions that have to be taken when designing evaluations. Stakeholders have to be consulted, programme 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. It was noted in Chapter 3 that it is good practice for those commissioning evaluation to leave scope for evaluators to specify their preferred method of 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.. Unfortunately this flexibility is not always allowed. 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: It was noted in Chapter 1 that 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, mid-term or intermediate stage once a programme is under way, and ex-post once it has been completed. 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).

The OECD Public management service *Guidelines on Best Practice for evaluation* suggests that the following considerations should be taken into account when choosing appropriate evaluation methods:

1. *Definition and Objectives of evaluation* - Evaluations are analytical assessments, addressing the results of public policies, organisations or programmes,, that emphasise the reliability and usefulness of findings. Their role is to improve information and reduce uncertainty; however, even evaluations based on rigorous methods rely significantly on judgements.
2. *Manage evaluation activities strategically:*

a. Organisation of evaluation should correspond to needs and priorities in different policy areas. It may be appropriate to systematise and institutionalise evaluations in key policy areas where the costs of collecting data is high and information limited. However, a more flexible approach will often produce better results and prevent evaluations from becoming paperwork exercises.

b. Development of evaluation skills in different organisations ensures the necessary range of evaluation methods and perspectives (e.g., drawing from both internal and external evaluators), and that each evaluation is designed in accordance with its unique set of issues related to objectives, focus, credibility and intended use.

3. *Enhance credibility* - Methodological quality of evaluation (encompassing issues such as relevant criteria, adequate data and evidence and reliable and clear findings) has an effect on the credibility of evaluation. Quality assurance and an open and frank dialogue can improve credibility by exposing and rectifying potential weaknesses in evaluations.

Source: OECD Public Management Service (May 1998): *Best Practice Guidelines for Evaluation – PUMA Policy Brief No. 5*

2) Quantitative versus Qualitative : A false debate?

A common-sense distinction is often made between quantitative and qualitative methods, techniques and data. 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'. Indeed 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 and procedures, 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 genuine 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, questionnaires and observations, 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 local and sectoral programmes. 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 2002 and then in 2006) 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 in the round 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 Sourcebook 2. 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[11] for gathering and analysing data are important in socio-economic development because:

- *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.

[11] See also the example on Improving the efficiency of the Indian Employment Assurance Scheme, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/boxes/indian_employment_assurance_en.htm

- *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 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 detailed and often 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 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 especially in circumstances of social exclusion and bottom-up development, 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. As has been suggested earlier, putting in place basic data systems should be part of the planning of programmes as well as evaluations.

Choices for different types of interventions

1) Policy and Sectoral priorities

Particular approaches to evaluation tend to be associated with particular policy areas. This applies equally to the choice of methods and techniques. In many ways this follows from the nature of the policy area concerned. In agriculture, within Europe, evaluations of the impacts of subsidy regimes on production would be a more common approach than when evaluating human resource interventions. In the latter case, ways of measuring the accumulation of human capital and labour market impacts would also come to the fore. There is an extent, therefore, to which particular methods and techniques are properly associated with particular policy areas. 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 extensive use of experimental methods.

In Sourcebook 1, a fuller range of policy priorities and themes are described in terms of the types of evaluation methods, data and indicators that are commonly used. It is necessary, nonetheless, to repeat the warning made earlier. There is a tendency for evaluators who work intensively in particular policy areas such as - environment, transport, human resources, science and technology, SME development - 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.

2) Local and territorial development

Socio-economic development has always had a strong territorial dimension. Over recent years local development has been seen as increasingly relevant in terms of realising the socio-economic potential.

It is not usual to be able to use econometric methods in the evaluation of socio-economic development at the level of a locality or small territorial unit, mainly because interventions and programmes tend to account for a relatively small proportion of net resources and many other inputs (transfer payments, other programmes, etc.) play a role.

Econometric models are appropriate where the programmes cover an entire territory, provided that the situation meets the following criteria:

- The funds spent in the framework of the programme are significant compared to the economy of the territory concerned, and
- The territory must be large enough (a country or a large region), or closed enough (an island) for the functioning of its economy to be considered in isolation.

When these conditions are met, the evaluation team can use several techniques borrowed from regional economics or macroeconomics. These include:

- Shift-share analysis consists of projecting national economic trends onto the economy of a region. This technique is used to estimate a policy-off situation. By comparing this with the policy-on situation, the global impact of the programme can be evaluated.
- The input-output model and econometric models are used to simulate the economic development of the region. These techniques are generally applied ex ante to two types of assumption (with / without the programme) and can provide an overall estimation of probable macro-economic impacts.

When an evaluation concerns interventions with a more defined scope, it is possible to carry out an in-depth analysis of the causal links between the intervention and its effects. Several techniques may be used in this context:

- Variance analysis, factor analysis and cluster analysis are used to reveal similarities and dissimilarities within a sample of observations, to create typologies, and to identify exogenous factors associated with the production of the impacts.
- The Delphi survey was designed for estimating impacts. It is well suited to ex ante evaluations that rely on secondary data. The technique mobilises and analyses data through the intervention of experts. It is therefore similar to the expert panel that is also suitable for data analysis.
- Control groups are used to estimate net effects by noting the difference between a group of beneficiaries and a group of non-beneficiaries.
- Regression analysis is used to estimate net effects and to determine whether the causal links between the intervention and its effects are statistically significant.
- Case studies, group interviews and participant observation are techniques for observation, but can also be used flexibly for content analysis and comparing and analysing data. It is possible to estimate effects (in the form of a range, from minimum to maximum effects) by means of case studies carried out on a selection of project level interventions.

Because local development starts from the analysis of local potential, capacity and needs, its evaluation is particularly suited to participatory methods that elicit from stakeholders and local citizens, their priorities, attitudes and behaviours. It is in these local development settings that the active engagement of local stakeholders in an evaluation including participatory, self-evaluation and empowerment orientated evaluations are most useful. These approaches are closely aligned with community development strategies, which are themselves often deployed in local development

settings. Of course, the analysis of current socio-economic baselines will be amenable to the traditional range of economic and statistical analyses. Furthermore, comparisons across areas, which are sometimes favoured (e.g. benchmarking), requires that standard indices and measures are applied in order to judge outputs and results.

Local development in particular is often understood as a process that, over time, moves through a number of stages and in which the consequence of each stage affects those that follow. For this reason, process evaluation methods that track development over time are especially useful. These might, for example, include: tracking critical incidents over time; encouraging local residents to keep diaries; creating panel studies (longitudinal surveys of the same respondents) which are revisited at different stages in the development process and evaluation reviews by local stakeholders (see local evaluation, Case studies, Participatory approach) .

One of the characteristics of local and territorial development is the importance they attribute to the interaction of many different factors that contribute to the development process or conversely are responsible for underdevelopment. For these reasons, evaluations of local and territorial development need to include methods that identify, describe and measure interactions between different interventions and the relative contributions that they make as well as their synergies (their matrix of cross impacts). Whilst this can sometimes be achieved using indicators, there is also a need to develop models that show the relationship between different factors and different interventions.

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. There are a range of such methods and techniques described in Sourcebook 2. These would include, for example, input-output analysis and efficiency analysis.
- 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. Comparisons can also be made with external examples through benchmarking. Where there is no easy way to compare externally, as is often the case in the context-specific world of socio-economic development, comparisons may be made on a before and after basis showing changes over time. 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 case material and qualitative descriptions of development outcomes to support more abstract descriptions in terms of finance or money alone.

- 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? From a positivist perspective, 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, has become more common in these kinds of evaluations. 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 as a large a number of evaluations as are available. As knowledge production has become more important with the commitment of policy makers to evidence-based policy-making, various kinds of meta-analysis have become widespread. This form of 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. Of course, approaches to institutional and network strengthening can be pursued in a much more direct way. For example, studies may be undertaken of the administrative capacity of particular partner organisations in order to help them adopt more suitable management processes and information systems.

Choices for different programme/policy stages

1) Formulation: Identifying needs and priorities

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, mid-term, 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 poses certain demands for evaluation in most major programmes:

At the *formulation* stage, there will be an emphasis on identifying needs and clarifying objectives;

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, intermediate outcomes 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.

Formulation: Identifying needs and priorities

Socio-economic development usually starts from two perspectives.: a positive perspective about the potential for development and a negative perspective about the needs and problems to be overcome. Methods will need to describe baseline circumstances probably in comparison with the circumstances of other eligible sites for development. In terms of hard measures, techniques such as benchmarking and the analysis of administrative data on income levels, qualifications, participation in the labour market and market characteristics (such as the proportion of economic activity devoted to out-region exports) will be necessary. Economic analyses of the profile of the territory or region can be useful not only to identify gaps in certain activities. They can also be used to reveal where there is potential for new activity, especially when the basis for comparison is another territory or region that shares some characteristics with where development is being planned. More sophisticated statistical and macro-economic models will also set up a comparison with comparable settings (territories, similar sectors perhaps in other countries) that can be used for explanatory purposes at a later stage in development.

Given the importance of stakeholders in planning and delivery of socio-economic development programmes, methods that can involve stakeholders at an early stage will be useful. These can range from consultative and participatory methods focus groups, local polls, public meetings, etc., through to more formal techniques such as SWOT analysis undertaken with different groups of stakeholders to elicit their understandings of what can be changed to what advantage.

Although much of the content of priority setting will come from the political system, there are also methods that can be used to deepen an understanding of what is needed including, for example, issue mapping or concept mapping which can provide a basis for identifying, grouping and prioritising potential interventions.

2) Design: Interventions and organisation

Policy and programme formulation usually entails the identification of starting circumstances and of desired goals and 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 goals. 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 including local citizens 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.

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.

3) 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. Indicators can be used as part of this tracking process.

4) 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 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. 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), and
- Indicators based on contextual data or administrative data provided by public authorities.

In local development programmes in particular, a participatory dimension will be an important part of evaluation methods. This is not simply to ensure that the voice of beneficiaries is included. It is also in order to ensure that local groups, citizens, trade associations etc are able to make their own judgements about the success of programmes. It is, after all, their judgements together with self-evaluation techniques - that will be important given the bottom-up orientation of these programmes, rather than the judgements of evaluators or policy makers based on more abstract criteria.

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, EUROSTAT[12] and other data bases kept by DG REGIO,
- 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. Typically, for example, the geographical unit for which administrative or statistical data is gathered does not conform with the boundaries of the socio-economic development programme in question.

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

- 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, but it is necessary. 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, e.g., through monitoring, over which the evaluation team will have less control. However even when data are collected directly by the evaluation team its strengths, limits, scope and relevance need to be though 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 socio-economic programme will be held in many different places. Simply mapping out where data is held

[12] http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL

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 for example 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:

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, e.g.,

- 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;
- 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 an 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. This will ensure that for example SME managers will see their evaluation questions included in the evaluation agenda and therefore see the results as relevant to them.
- 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 as a whole. 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 and often what begins as qualitative information is transformed into quantitative data through various analytic methods. 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 in terms of data. 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 the finer details of experience need to be described. The choice between such strategies must ultimately depend on what questions an evaluation is trying to answer. Satisfaction surveys among programme beneficiaries and participatory evaluations among excluded groups will use opinion data quite differently.

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 goal, objective and/or target. (Indeed, indicators are most helpful when objectives have been specified in terms of targets or milestones that apply 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 and some result indicators where data are provided directly by the projects, sample checks should verify the

data. For impact indicators, it is probably better that data are measured 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.

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. This is, however, a necessary but not sufficient quality. The following are examples of indicators that are readily understood: rate of budget absorption; percentage of regional firms assisted; number of net jobs created; 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, generic and core indicators
- In relation to the scope of information: Context and programme indicators
- In relation to the phases of completion of the programme: Resource, output, result and impact indicators
- In relation to evaluation criteria: Relevance, efficiency, effectiveness and performance indicators
- In relation to the mode of quantification and use of the information: Monitoring and evaluation indicators

The most useful of these typologies for socio-economic programmes is the distinction between: resources; outputs; results and impact indicators. Contextual indicators, which can be the same as impact indicators in some cases, is a further useful category.

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; progress rate of the building of a road; hectares of urban wasteland rehabilitated; capacity of purification plants built; number of trainees whose training was paid by the programme; and percentage of this training of which the quality is certified.

Result indicators represent the immediate advantages of the programme (or, exceptionally, the immediate disadvantages) for the direct beneficiaries. An advantage is immediate if it appears while the beneficiary is directly in contact with the programme. The full results may be observed when the operator has concluded the action and closed off the payments. Result indicators are generally easily known to the operators, so they are generally quantified during monitoring.

Result indicators provide information on changes which occur for direct beneficiaries, for example, time saved by users of a road; reduced rates for telephone calls; qualifications earned by trainees; new tourist activity generated by a farmer; use of new productive capacity created by a firm; and the satisfaction of businesses which have received consultancy services.

It is at the time that beneficiaries receive support or programme services that results can be quantified. Either direct measurements are made (e.g., by counting the number of trainees recruited during their training) or the direct beneficiaries are asked to state the advantages they have obtained (e.g., by means of a questionnaire at the end of a consultancy mission).

Impact indicators represent the consequences of the programme beyond its direct and immediate interaction with the beneficiaries. An initial category of impacts group together the consequences for direct beneficiaries of the programme, which appear or which last into the medium term (specific impacts), e.g., traffic on a road one year after it is opened; the placement rate of trainees after twelve months; sustainable jobs created in an industrial plant built with programme support; and the survival rate of businesses created with programme support. Some impacts are unanticipated (spin-offs) but indicators are rarely created for unanticipated impacts.

A second category of impacts consists of all the consequences that affect, in the short or medium term, people or organisations that are not direct beneficiaries. These impacts may be similar (e.g., improvement of the quality of life for people living near a rehabilitated industrial wasteland; improvement in the quality of beaches near a new purification plant). They may, in contrast, spill over to affect people or organisations far from the programme, as in the case of macro-economic impacts.

The mechanisms of impact propagation can be separated into two categories: market effects (e.g., impact on suppliers or sub-contractors of the assisted firms) and non-market effects (e.g., positive impact of the improved image of the region or negative impact of deterioration in the environment). Because non-market effects or externalities are not reflected in the price system on which individual

socio-economic actors largely base their private decisions and because these decisions have economic consequences for other actors, it is particularly useful to take these effects into account in the context of a public programme.

Because of the time lag or their indirect nature, impacts cannot easily be known to operators during their daily management of the programme. Impact indicators are therefore quantified from time to time only, usually during evaluations. One way of establishing impacts is to carry out a survey of direct beneficiaries, for example a year after they have left the programme. The questions asked might concern facts (e.g., how many new jobs have been created since the support was obtained?) or opinions (e.g., how many jobs would have been lost without the support?). When analysing values for impact indicators, evaluators may pay particular attention to causality. To what extent do those values relate to the programme or intervention being evaluated and what other factors may have had an influence?

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 effectiveness.
- 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 unemployment rate or disparities between infrastructures often play a decisive role.

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. Indicators depend on quantification and are also very useful for clarifying objectives. The Box below provides an example.

Defining indicators can clarify objectives:

Strategies for supporting the competitiveness of SMEs often include the provision of consultancy services. The objectives of these measures may be expressed in terms of the number of businesses to receive particular types of consultancy services. The indicator serves not only to quantify the objective, but also to define the expected service. For example, the definition of receipt of a service by a SME might only include significant consultancy missions amounting to more than 5 days of consulting.

Once defined and adopted, the programme is implemented. It is monitored and evaluated on an on going basis. At this stage indicators are indispensable for circulating, in a simple and condensed form, information required by programme managers. Typically, 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 rate of satisfaction of beneficiaries, the number of jobs created.

The programming cycle ends with an ex post evaluation, of which one of the main functions is to report on the programme results and on the extent to which aims have been achieved. The use of indicators is strongly recommended at this stage in so far as it allows the communication of simple information that is immediately understood by a wide public, e.g., cost per job created or rate of placement of jobless people assisted.

Indicators for integrated programmes

Most socio-economic programmes adopt integrated strategies, in other words, they try to solve all the problems affecting a given territory and they use all available instruments for intervening in that territory. This characteristic necessarily entails a multiplication of needs for indicators, which would lead to confusion if the programmes were not highly structured. Programmes financed by European Cohesion Policy are usually structured on three levels:

- the overall programme level to which the global objective is related, for example, economic development or employment. This level consists of a small number of priority axes (usually less than six) which break down the global objective into its main strategic dimensions;
- the measure level (from one to several dozen) existed up to 2006 and corresponded to the basic unit of programme management. Each measure had its own specific management apparatus. While measures are no longer part of the Cohesion Policy programme per se, managing authorities continue to operate "measures" for the purposes of managing their programmes;
- the project level (often many hundred), which is the implementation unit of the programme, since each project is a point of articulation between the programme and its beneficiaries.

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.

A team should then be appointed to support the group and provide a secretariat. Typically the team members belong to the authority managing the programme. They should have the required human and financial resources. The team must, in particular, ensure that the system of indicators clearly reflects the programme objectives and favours comparability. It is preferable for the same team that is responsible for creating indicators to subsequently be responsible for the implementation of the system.

The public may also be involved in designing the system of indicators. An example of involving beneficiaries in the choice of indicators from American experience (Benton Harbour region, see the Box below), started with a series of focus groups involving representatives of regional enterprises. The work of these groups made it possible to select indicators most likely to attract the public's attention and to be understood by it.

Involving beneficiaries in the choice of indicators:

Political authorities in the Benton Harbor region (USA) set up a system of context indicators with a view to assessing the economic development of the Berrien metropolitan area. The key elements in the description of economic development were based on a series of focus group interviews involving the leading entrepreneurs in the region. The indicators chosen were, for example:- For the availability of qualified human capital: spending on education per child; the teacher-student ratio; the school dropout rate.- For the growth and diversification of the economy: per capita income; employment rate in the industrial sector; value added in the trade sector; index of sectoral diversity; rate of employment in SMEs; value and number of residences that were issued building permits.- For the quality of life: the cost of living (relative), the rate of property crimes and crimes against the person.

Erickcek G.A. (1996) The Benton Harbor Area Benchmarking Data System, Michigan: WE Upjohn Institute

The main suppliers of information are the operators who implement the programme 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.

It is also advisable to involve the operators in a preliminary test of the system of indicators. The recommended procedure starts with the selection of a few volunteer operators who will participate in the design of the system. These volunteers should represent all the components of the programme. They help to choose the indicators, to define them and to plan the information collection process. They express their needs in terms of information feedback (frequency and form of information fed back to them). The test comprises an initial quantification of all the indicators by voluntary operators. The normal duration of such a test is a year. Depending on the conclusions of the test, and after introducing the necessary modifications, the system is validated. The definitions and the data collection and restitution procedures are clearly established, and a manual is written.

Information relating to the context is drawn from statistics. It is therefore advisable to involve an expert with recent and complete knowledge of exploitable statistical data, in designing the system. Depending on the case, this expert will belong to a statistics institute or a university or, if possible, to the institution running the programme.

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 therefore 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 the 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.

In several European regions, evaluations have shown that a few dozen indicators are enough to meet the information needs of the authorities running the programme (as in Northern Ireland, see Box below). This does not mean, however, that additional indicators may not be required to meet the operators' information needs.

The recommendation of an evaluation: reduce the number of indicators from 330 to 52:

A socio-economic programme was financed by the ESF for the period 1994-99 in Northern Ireland. An intermediate evaluation was made of this programme, by synthesising six separate evaluations of sub-programmes. The programme designers had chosen 330 indicators for the monitoring. These indicators had been grouped into a single data base that proved to be difficult to use and created problems of manipulation and availability of information. These problems considerably reduced the use of the indicators. The evaluation team recommended a solution consisting of:

- **Choosing a number of context indicators situated between the macro-economic indicators and the sub-programme indicators. These indicators, intended to reflect the global impact of the programme, are divided into three categories: economic growth, internal cohesion and external cohesion.**
- **Choosing a small number of programme indicators by limiting them to the main results and impacts;**
- **Delegating the quantification and use of the other indicators to the operators.**

In this way, the size of the system would be reduced to 52 indicators directly related to the main objectives of the programme. The recommendations were successfully applied.

Source: Colin Stutt Consulting, (1997) Northern Ireland Single Programme 1994-99; Mid Term Review External Evaluation

Suggestions for limiting the size of systems of indicators

Suggestions for limiting the size of systems of indicators are typically based on the use of generic or core indicators or on the grouping of indicators by category of beneficiary. A lighter system limits the collection and circulation of information to the most essential elements at the programme level. On the other hand, this means that the progress and results of each action will not be monitored in a detailed and centralised manner. It also means that the system focuses less on the decisions to be made by the operators and more on those to be made by the authorities managing the programme.

Core indicators

At the stage of the preparations for the update of the Mid-Term Evaluation of 2005, the European Commission introduced the concept of "core indicators" in the relevant guidance document. The idea was to identify some basic (output and result) indicators which could be aggregated and compared (with caution) across regions and Member States. The core indicators were revised and developed in the Commission guidance on indicators for the 2007-2013 period. While urging the use of core indicators, the guidance paper also stresses the need for the responsible authorities to develop better quality result indicators, in line with programme objectives and with the stronger strategic focus of the new programming period.

4) Assessing quality

The use of indicators will be far greater if their quality is constantly improved. Evaluation has an important role to play in assessing the quality of systems of indicators and recommending ways to enhance them. Although there is no standard method for this quality control, an approach is proposed based on the following criteria, which are divided into two groups: quality criteria applicable to each indicator and quality criteria applicable to the entire system.

Quality criteria applicable to each indicator

The first quality criterion for an indicator is the capacity for it to be quantified at regular intervals. Sometimes one or more indicators featured in the programming documents have never been quantified and therefore cannot inform the evaluation or the monitoring of the programme. The availability of data to allow quantification is the primary factor to be considered. Monitoring indicators (inputs, outputs and some results) should be quantified at each monitoring meeting, that is to say, every six to twelve months. Evaluation indicators (some results and impacts as well as context indicators) are quantified less frequently, typically annually or every three to six years.

Once an indicator has been quantified, it may take several months or even years before the information can really be used for monitoring and evaluation. This is particularly true for certain context indicators drawn from statistical publications. The freshness of information is an important quality criterion. Sometimes statistics are published two years or more after the collection of the data.

When evaluating programme effects, the indicators chosen must be such that the programme is capable of bringing about a change in the indicator value. The capacity for interventions to impact on an indicator is known as sensitivity. Take the example of an intervention supporting exports; the

turnover of businesses is not a sufficiently sensitive indicator. A better indicator would be the turnover relating only to new customers contacted with the support of the programme.

The results produced by applying the indicators need to be reliable and credible. Reliability tends to apply to facts and figures and can be defined as the fact that the same measurement, taken by two different people under identical conditions, will produce the same value for the indicator. In cases where indicators are quantified on the basis of questions put by one person to another, reliability can no longer be defined so mechanically, although the tests need to be credible. Credibility tends to depend on the soundness of the method, although the independence and reputation of the evaluation team may also be important.

The usefulness of an indicator depends largely on whether it allows for internal comparisons between different measures of the programme or inter-regional external comparisons. The comparability of the indicator is therefore a quality criterion.

A further quality criterion of an indicator is normativity. Indicators should relate to outcomes that can be judged to be satisfactory or not. Indicators should avoid ambiguity. Any indicator value must therefore be compared to a norm, for example: objective to be met; norm to be surpassed; or European average to be attained.

A good indicator must be understood by everyone who has to use it. In the minds of both decision-makers and the public, the meaning of the indicator must be the same as for the programme managers and the project promoters providing the source data. It must accurately reflect the concept to be measured. This is sometimes referred to as validity.

Quality criteria applicable to the entire indicator system

The following criteria are proposed to assess indicator systems:

- The indicators selected should cover a sufficiently large proportion of the programme measures. This coverage should be equal to or greater than three-quarters of the planned expenditure.
- The system should consist of a good balance between indicators in the different categories.
- The system of indicators should be simple. The selectivity criterion requires that the programme managers' capacity to absorb information be respected. The information must therefore be limited to a maximum of a few dozen indicators.
- The relevance of the system implies that the indicators are developed primarily for those measures or themes that have significant implications in terms of decision-making. For example, measures with a very high budget; innovative measures; themes considered to be strategic.

Very often the setting up of indicators will not start from scratch and wherever possible systems and indicators should be consistent with those already operating.

5) Applying indicators

Using indicators to make comparisons between programmes

Using indicators to make valid comparisons between programmes is important but difficult. This is due to various factors, such as the diversity of interventions within a programme, the diversity of regional contexts, or the incompatibility of definitions. For example, depending on the regions and programmes, tourist trips may be counted in terms of the number of visits or the number of nights stayed; trainees may be counted in terms of the number of participants in training course or in hours of training provided; and environmental protection may be measured in terms of the number of projects, the number of sites or the number of hectares protected.

Comparability may be sought and obtained through exchanges between managers in different regions or countries.

Comparability is often easier to obtain and more useful at all levels if it results from a co-ordinated collective effort at a higher geographical level. This approach has the advantage of multiplying the possibilities for comparison, and also allowing for the aggregation of indicators at the regional or national level. The Scottish Office created a series of standard indicators applicable to seven Scottish programmes (see Box below). Although mainly concerned with contextual indicators rather than evaluation, the Urban Audit provides an example of efforts to achieve comparability.

A set of standard indicators for several programmes

Within the framework of its responsibility for monitoring and evaluating seven programmes co-financed by the European Union, the *Scottish Executive* developed a series of standard indicators. Some of these indicators concern the context: rate of employment and unemployment; productivity; workforce; average income; etc. Some output and result indicators have also been standardised, as shown by the following examples, applicable to the infrastructure measures for enterprises:

- **New buildings built (square metres)**
- **Buildings renovated (square metres)**
- **Development of new sites (hectares)**
- **Improvement of existing sites (hectares)**
- **New / improved road access (kilometres)**
- **Surface area for which the access roads were built or improved (hectares)**
- **Rate of occupation of the new buildings (percentage after one year/percentage after three years)**
- **Number of training places created (number)**

To obtain a high level of comparability, the Scottish programme managers organised several meetings attended by representatives of all the programmes. Standardisation was the result of a long process of collective discussion.

Public communication

Systems of indicators should be useful for decision-making. They are also important for accountability purposes, for example to the European or national parliaments, to regional or local elected representatives, to socio-economic partners, to journalists and, through them, to citizens and taxpayers.

If systems of indicators are to serve as a basis for public communication, a small number of indicators that can immediately be understood by lay people must be selected, quantified and published. The publication of such indicators is normally organised in the form of simple tables with accompanying commentary, for example in an annual review. More detailed information can also be made available through an "observatory" open to the public, or through a system of consultation on the Internet.

In defining these publicly accessible indicators, priority should be given to generic indicators (applicable to many different actions within the same programme) and standard indicators (allowing for comparisons between programmes in different regions or countries). Moreover, these indicators should be understood by all the partners without long explanations and without any misinterpretation of their meaning.

Using indicators to improve management

1) Managing performance rather than resources

Increasing numbers of systems of indicators are created for the purposes of "performance management". These systems are a form of New Public Management that emphasises results and impacts obtained, as opposed to older forms of management based on the allocation of resources and the control of outputs.

Managing performance rather than resources

In the spirit of performance management, operators are endowed with greater autonomy in the use of their resources. In return, they commit themselves to clearer objectives as regards the results and impacts to be obtained. They have to measure their performance in order to evaluate themselves and submit periodic reports. This new balance between decentralisation and performance measurement is at the base of many developments in public administration[13] .

In many European regions, the administrative culture has remained impervious to decentralisation and to performance management and the development of result and impact indicators is generally considered to be difficult. Programme managers are more familiar with resource and output indicators. Cultural changes are, however, slowly but surely taking place in certain countries, under pressure from administrative reforms initiated by national governments. The monitoring and evaluation of programmes co-financed by the European Union has been a factor encouraging performance management in terms of results and impacts.

[13] For example, the US Government Performance and Results Act (GPRA) of 1993 made the use of new indicators compulsory through the entire government administration.

2) Interpreting and comparing indicators

Situations exist in which indicators speak for themselves, but these are exceptions. In general, indicators have to be interpreted by means of the relevant comparison or breakdown. In one example, the comparison of three indicators showed that the training financed with EU funding did not reach the long-term unemployed as it should have (see Box below).

Comparing several indicators to reveal a phenomenon:

Within the framework of its intermediate evaluation of the ESF in the Italian Objective 1 regions, the ISFOL gathered and interpreted a series of monitoring indicators. In certain cases, the comparisons made between these indicators led to strong and clear conclusions. For example, the part of the programme benefiting the long-term unemployed can be measured by means of three indicators:

- **planned expenditure on the long-term unemployed: 18% of the total planned expenditure**
- **funds committed to the long-term unemployed: 11% of the total funds committed**
- **funds actually spent for the long-term unemployed: 8% of the total funds spent.**

The comparison of these figures clearly shows that the long-term unemployed were successively disqualified at each step of the programmes implementation. For a reader who glances at the table of indicators and sees these three figures in isolation, the information would be meaningless. To identify the problem of the funding not reaching the long-term unemployed, it was necessary to break down these three indicators in terms of the length of unemployment of the direct beneficiary, and for the ISFOL to compare the three pieces of information alongside each other.

Whenever possible, it is useful to compare programme indicators (for example, km of roads (motorways) built) with the appropriate context indicators (for example, baseline km of motorway in the region). This type of information can be presented in a concise way by means of derived indicators that express the programme outputs or effects as a percentage of the context indicator.

In order to be useful in evaluation work indicators need to be used in conjunction with qualitative findings. To interpret indicators, it is necessary to consider the context as a whole, the factors which help to facilitate or hinder the performance of the programme, the rationales of the programme, and the process of implementation.

One technique consists of asking an expert panel, to examine the combined quantitative and qualitative elements of the situation, to interpret the performance measures (see Box below).

Expert panels to assess University Research Centre performance:

Expert Panels have been used for some time to evaluate the performance of university research centres in the UK. The work of all the UK research centres is monitored quantitatively and periodically evaluated by expert panels. In their evaluation, the experts take into account indicators (e.g. number of publications) but also purely qualitative elements (e.g. quality of each researcher's four best publications). The evaluation results in a classification of the research teams (on a scale of 1 to 5 now a star rating). The allocation of public funds is directly related to this classification (Bone & Carter, 1997)

The interpretation of indicators often makes use of comparisons between the performance of two interventions. This type of comparison mainly concerns efficiency indicators, for example: cost of a trainee's training; cost of a job created; cost of a home connected to the sewerage system. However, comparisons are not always acceptable for programme operators and managers. In particular, performance comparisons will seem unfair to managers who work in the more difficult contexts. To make comparisons acceptable, the comparison procedures must be considerably refined with a view to taking account of the additional challenges of interventions taking place in more difficult environments.

An example of how this issue was addressed in the education domain is given in the Box below. The technique consists of reconstituting, for each intervention to be compared, a group of several similar interventions implemented in other regions in an environment with a similar degree of difficulty. Comparisons were made with the average of the interventions of this "benchmark group".

Example of comparison of performance accepted by the operators:

The comparison of performance between training organisations often stumbles over the fact that the different groups of beneficiaries do not have the same difficulties. The question is how to ensure one is comparing like with like. A few years ago an evaluation team had to compare 133 school districts in different states in the U.S. The team decided to compare each school district to a "benchmark group" of 14 other school districts with similar pupil characteristics (poverty; home language; parents' level of education; belonging to a minority group; etc.). To form the standard groups, the evaluation combined a factor analysis and expert panels. At the end of the evaluation, the validity of the performance comparisons were accepted without question.

Henry et al, (1992)

3) 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 in the Box 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.

Adverse affects of a performance management indicator system: 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 to be 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.

Creating performance incentives

There are several ways of using indicators to promote an improvement in operators' performance. These include:

- Operators with poor performance receive specific technical assistance to help them progress. If the situation does not improve, the budget is restricted. This method works on the principle that it is not the mistake that must be penalised but rather the inability to correct mistakes.
- Operators with the best performance are granted greater autonomy and are controlled less.
- Operators with the best performance receive support for presenting their outputs and results to the general public.
- Operators who did not perform well enough are disqualified from the selection procedures for future projects (and example of this is given in the Box below).
- Operators with the best performance are offered additional funds.

Selection of projects in terms of their effectiveness:

Within the framework of the Objective 2 programme for the period 1989-93 in the Pays-de-la-Loire region (F), the ESF actions were managed by the Regional Council. For the management of the ESF, they applied a project selection method based on performance management.

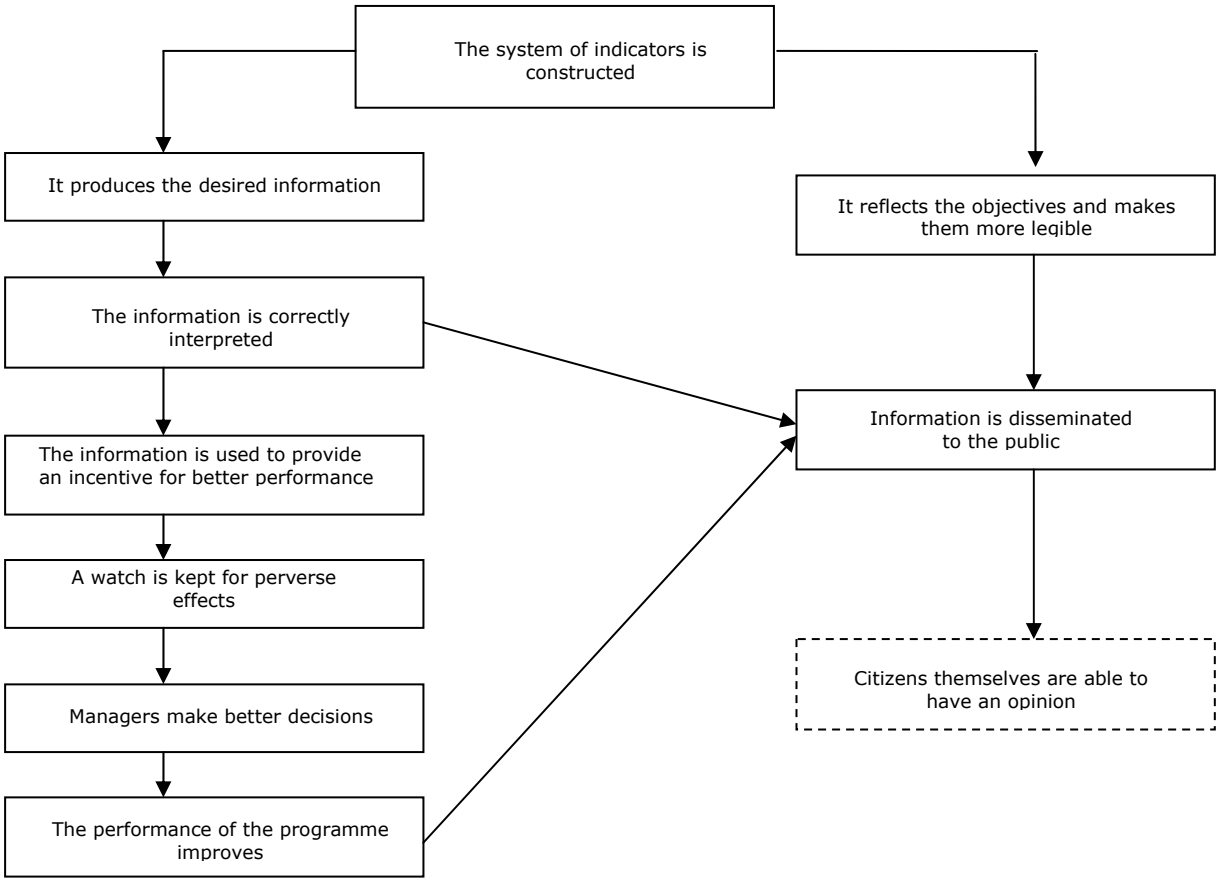
The monitoring system obliges all organisations that offer subsidised training to provide the Regional Council with an evaluation of the placement rate of its trainees after six months. Secodnary/additional applications for funds are granted first to those organisations with the best placement rates.The effect of this mechanism is to favour those organisations which are most tuned in to the needs of business and which have the best impacts.

4) The usefulness of a system of indicators

The figure below summarises the main messages elaborated above. It presents a set of assumptions that are made, often implicitly, when a system of indicators is created with the intention of improving a programme's performance.

The presentation emphasises that the construction of a system of indicators is not really a first step. At least six other steps occur before the system is actually useful.

A system of indicators:



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 always 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 socio-economic development 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. The participatory nature of local development building on the potential and ambitions of local stakeholders and citizens is especially suitable for both qualitative methods and participatory methods.
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. It is no good investing in sophisticated methods to gather data but to be relatively simplistic in the way data is analysed.
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. At the end of the day 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. In socio-economic development programmes in particular there are a host of problems to be resolved. Different partners have to be willing to share

information, excluded groups often distrust evaluation as one further example of official behaviour and need to be brought on board, 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 evaluation as a whole.

10. The quantitative/qualitative divide is overstated. Data are often more of a continuum, beginning life as qualitative and once analysed becoming quantitative. Weaker forms of quantitative data (e.g., categorisations or ranking) are close to qualitative data. What is needed when evaluating socio-economic development programmes is qualitative data able to capture subtleties, people's experience and judgements and quantitative data to provide overviews, for example to aggregate results of an intervention and provide a comparative perspective.
11. Well conceived indicators and monitoring systems are a powerful adjunct to evaluation. Very often evaluators depend on monitoring systems which are indicator based. If these are not put in place early in the programme design cycle it may be too late to create such systems later on.
12. Over elaborate indicator systems may be counter productive. Whilst there is a temptation in multi-sectoral and multi-objective programmes to measure everything, this should be resisted. This can be costly and the results difficult to use.
13. Indicators are often used for management and accountability purposes. It can be difficult to reuse indicators that have been exclusively developed for such purposes as part of an evaluation. There can be too much pressure to shape information in positive ways or at the very least make sure that bad news does not come through. On the other hand, within a well-developed performance management culture these kinds of indicators can help improve programme content and management.

Annex A - The main stages of evaluation

Ex-ante evaluation

Ex-ante evaluation takes place at the beginning of the cycle before a programme has been adopted.

This form of evaluation helps to ensure that the programme is as relevant and coherent as possible. Its conclusions are intended to be integrated into the programme when decisions are taken.

Ex ante evaluation focuses primarily on an analysis of the strengths, weaknesses and potential of the Member State, region or sector concerned. It provides the relevant authorities with a prior judgement on whether development issues have been diagnosed correctly, whether the strategy and objectives proposed are relevant, whether there is incoherence in relation to Community policies and guidelines, whether the expected impacts are realistic, and so on. It also provides the required foundations for monitoring and for future evaluations, by ensuring that there are explicit and, where possible, quantified objectives. It helps to specify selection criteria for the selection of projects and to ensure that Community priorities are respected. Finally, it helps to ensure the transparency of decisions by allowing for a clear explanation of choices made and their expected effects.

Ex ante evaluations are performed at the time when public authorities are involved in discussions and negotiations on the future programme. They are therefore subjected to strong constraints: pressure of deadlines, vague formalisation of the proposed programme to be evaluated, amendments to this proposal while the work is underway, demands for confidentiality, etc. The evaluation team must therefore be able to intervene flexibly and rapidly and be able to apply techniques for analysing needs and simulating socio-economic effects.

See also examples of ex ante evaluations in the Czech Republic[14].

Mid-term and ongoing evaluation

Mid-term evaluation is performed during the second stage of the programming cycle, during the implementation of the interventions.

Depending on the conclusions of mid-term evaluation, adjustments may be made during the cycle. This evaluation critically analyses the first outputs and results of interventions. It also assesses the financial management of the programme and the quality of the monitoring and of its implementation. It shows how and whether original intentions have been carried out and, where relevant, checks whether de facto changes have been made to the initial objectives. By comparison with the initial situation, it highlights changes in the general economic and social context and judges whether the objectives remain relevant. Mid-term evaluation also examines whether the evolution of Community priorities and policies poses a problem of coherence, and helps to prepare adjustments and reprogramming, and to argue them in a transparent manner.

[14] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/boxes/exante_nsrf_cz_en.htm

Mid-term evaluation relies heavily on information drawn from the monitoring system, but also on ex ante evaluation and on information on the context and its evolution. It generally consists of short and exhaustive exercises focusing primarily on the results of the programme evaluated, without attempting an in-depth analysis of impacts that have not yet had the time to emerge. It is, however, possible and advisable to refer to in-depth or thematic evaluations of former programmes when such analyses do exist. Mid-term evaluation has a "formative" nature, that is to say, it produces direct feedback into the programme that it is helping to improve as far as its management is concerned.

For the 2007-2013 period, mid term evaluation has been replaced by ongoing evaluation, preferably based on a multi-annual evaluation plan. This new approach aims to overcome the rigidities of the 2003 mid term evaluation and facilitate the evaluation of parts or themes across programmes at a time when those responsible for the programme have a need for information, evidence, analysis or judgements from evaluators.

An evaluation plan should, ideally, outline responsibilities for evaluation, the budget, an indicative list of evaluations to be carried out (which can be reviewed and updated periodically), the mechanisms for designing and managing evaluations (tendering, steering groups, etc.), and arrangements for using evaluations (dissemination, responding to recommendations, monitoring the implementation of agreed recommendations, etc.). Evaluations plans can be established nationally or at the level of an individual programme (either regional or sectoral).

The concept of ongoing evaluation is extremely flexible. By leaving it entirely to the discretion of the programme authorities to decide what should be evaluated and when, evaluation should become more of a management tool to help the performance of the programme. Programme authorities can, of course, decide to undertake mid term evaluations of entire programmes, as outlined above. However, given that entire programmes need no longer to be evaluated at fixed points in time, there is also an opportunity to carry out in-depth evaluations of particular areas of intervention, perhaps using more complex and rigorous methods. An ongoing programme of evaluation should support public authorities to find out "what works" in their programmes and what doesn't and to take remedial action where necessary.

Ex post evaluation

Ex post evaluation recapitulates and judges the entire programme, particularly its impacts. Its aim is to account for the use of resources and to report on the effectiveness and efficiency of interventions and the extent to which expected effects were achieved. It focuses on factors of success or failure, and on the sustainability of results and impacts. It tries to draw conclusions that can be generalised and applied to other programmes or regions.

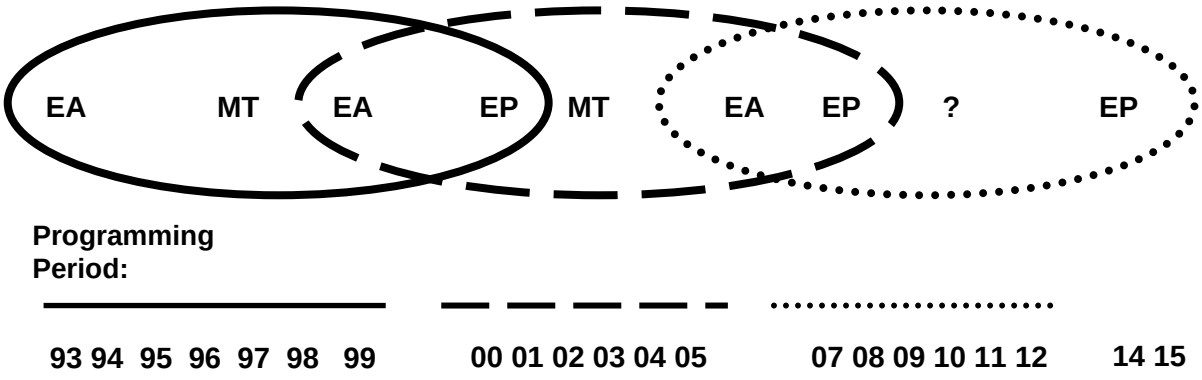
Ideally, the results of this evaluation should be available when the next programme is planned, that is, at least a year before the end of the programme. However, for the impacts to have been produced, ex post evaluation would have to be performed two to three years after the end of the programming period. While waiting for this period to pass, a provisional review is often requested shortly before the end of the programming cycle, in liaison with the ex ante evaluation of the following cycle.

Impact analysis is always a large-scale exercise if performed systematically. Ex post evaluations therefore tend to involve surveys in the field and to take place over long periods lasting from twelve to eighteen months.

Successive programming cycles

The sequence of evaluation phases in successive cycles creates overlaps that have to be organised as efficiently as possible to avoid duplication. The basic principle is that evaluation during a programme should use evidence from evaluation performed on the preceding programme. The relative continuity of actions programmed from one period to the next makes it possible to use evidence from the recent past to judge the relevance of the new measures proposed. The following diagram shows that interactions are possible between evaluations performed at the different phases of several successive programmes.

Articulation of evaluation programming cycles:



Key: EA: ex ante, MT: mid-term, EP: ex post

Thus, an ex ante evaluation that prepares the adoption of a future programme has to take advantage of the results of earlier work, i.e.:

- the intermediate evaluation of the period drawing to a close. This evaluation will have produced conclusions on the first years of activity and on the ensuing programmes. It may have been completed by a final review of the outputs and results of the current programme, based on information from the monitoring system. Evaluations may also have been done on parts of programmes.
- ex post evaluation of the period preceding the current period, possibly completed by thematic evaluations and in-depth analyses. These evaluations will have made it possible to observe and analyse the impacts of former interventions which are similar to the planned interventions and which took place in a partly similar context.

Since the role of intermediate evaluation (mid-term or ongoing) is to check whether the objectives are still relevant and in the process of being achieved, it will be necessary to refer primarily to the monitoring system data but also:

- to the ex-ante evaluation and particularly to the diagnosis made in relation to the prevailing socio-economic context before the start of the programme, and which needs to be updated;
- to the ex-post evaluation of the preceding programme, of which the conclusions concerning the same areas of intervention could serve as references.

Ex post evaluation is based on management and monitoring data and on surveys in the field which will help to observe and analyse the real and sustainable impacts of the interventions. It refers to ex ante evaluation in so far as it has to report on the attainment of objectives. It refers to intermediate evaluation, particularly to identify the success or failures which were identified at that stage.

In so far as evaluation must draw conclusions from the experience of preceding programmes to improve future programmes, an interesting solution is to establish a multi-annual evaluation plan as recommended by the European Commission for the 2007-2013 period (and for the new Member States in 2004-2006). The idea is to identify the different possible evaluations and to establish their timing and content in relation to the political schedules and deadlines for decision-making of the evaluation "customers" at various levels.

GUIDE Annex B - Changes in Structural Fund regulations

1988-1993

At the Commission, a demand for transparency vis à vis taxpayers, on the sound use of funds from the Member States, has progressively taken hold. This aspiration is reflected in a simple expression: "striving for effectiveness in European spending".

The Single European Act, adopted at the beginning of 1986, introduced into the EEC Treaty a new section V containing article 130D. This article announced the reform of the Structural Funds, intended to improve their effectiveness with a view to enhancing economic and social cohesion in the Community. Similarly, in February 1992, the Maastricht Treaty included this imperative need for cohesion, and Article 130B stipulates that the Council is to define measures required for ensuring that the Funds are effectively used.

In accordance with these requirements, the regulations relating to Structural Funds (those of 1988 for the first generation of Structural Funds and those of 1993 for the next generation) included specific articles on evaluation, particularly Article 6 of the framework regulation and Article 26 of the co-ordination regulation.

Article 6, Paragraph 2, of the 1988 regulations stipulated that "in order to assess the effectiveness of structural interventions, the impacts of Community actions must be evaluated ex ante and ex post. For this purpose, applications for support by the competent authorities must include the necessary information so that it can be evaluated by the Commission".

These first regulations, adopted by the Council on 19 December 1988, concerned regional development plans which had to be submitted by 1 January 1989. For many national and regional administrations, the need to establish multi-annual programming was in itself a kind of revolution. Moreover, most competent authorities and the Commission services did not have the required evaluation skills, at least as far as evaluating Structural Funds was concerned. Consequently, evaluations carried out during the 1989-93 period were not of a very high quality, despite efforts by many actors, especially at the Commission. It was observed, in particular, that the framework defined by this first version of the regulations failed to create adequate conditions for the ex ante evaluation phase to take place in the normal way.

1994-1999

Acknowledging this partial failure, at the instigation of the Commission but also through the strong impetus given by certain Member States (the Netherlands and the U.K. in particular), the new July 1993 regulations considerably strengthened requirements concerning evaluation.

This new regulation (Article 6 of the framework regulation) specifies the notions of prior assessment, monitoring and ex post evaluation. The Articles detailing the content of programming documents submitted by the Member States, introduce the notion of "specific quantified objectives, where possible". Similarly, Article 26 of the coordination regulations presents a precondition clause

specifying that "support will be granted when a prior assessment has shown the socio-economic advantages to be gained in the medium-term, in relation to the resources mobilised".

In practice, this very strong requirement for ex ante evaluation, a Community variant of the British notion of Value-For-Money, was not applied very strictly. By contrast, the introduction of mid-term evaluation which had not existed before was successful. This strengthening of European legislation as regards evaluation was clearly a strong incentive for spreading the idea that evaluation was unavoidable if Community Funds were to be obtained.

2000-2006

With the third generation of Structural Funds the assessment of their effectiveness was reinforced. Community actions were henceforth to be the object of ex ante evaluation by the Member States themselves, of mid-term evaluation by programme managers, and of ex post evaluation, at the initiative of the Commission, for assessing their effects in relation to the objectives (their particular objectives and cohesion objectives) and for analysing their impact on specific structural problems.

The effectiveness of the Funds was to be measured at three levels: the overall effect on the objectives of Article 130 A of the Treaty (particularly the strengthening of social and economic cohesion), the effect on the priorities proposed in the Plans and provided for in each CSF, and the effect on the specific priorities selected for the interventions. Complementary evaluations, where relevant of a thematic nature, could be launched with a view to identifying experiences that are transferable from one programme to another.

In order to improve transparency, evaluation reports were to be made available to the public more extensively than in the past.

2007-2013

In terms of evaluation, the new Structural Funds organisation for the period 2007-2013 represents a shift from a mandatory approach to a more flexible on-going approach.

Ex-ante evaluation of all the operational programmes under the Convergence Objective (prior Objective 1) remains an obligation, while Operational Programmes under the Regional Competitiveness Objective (prior Objective 2) and the Employment Objective (prior Objective 3) should be evaluated either per programme, per Fund or per priority axis. Operational Programmes under the Territorial Co-operation Objective should be evaluated individually or in groups. The ex-ante evaluation of the new National Strategic Reference Frameworks (which replace the National Development Plan and CSF) is not binding but it is strongly recommended by the European Commission, in particular for the large Convergence Objective Member States.

Building on the mid-term evaluation experience from the previous period, which showed that the imposed deadlines were rather too rigid, the requirements too broad and the terms of reference insufficiently adapted to specific needs, the new on-going evaluation should be more driven by particular country or regional needs than by Regulations. It is anticipated that Member States will carry-out on-going evaluation linked to the monitoring of Operational Programmes but this is only

required in cases of significant departures from initial programme goals and of proposals for programme revisions.

Overall, the Structural Fund Programmes for 2007-2013 should be more strategic, focused on longer term global objectives, which will be consequently translated by the Member States into more operational and short-term objectives and measures. Accordingly, evaluation should also be more focused on country contexts and on meeting the needs of decision-makers as well as of the general public.

Links:

Regulation:

http://ec.europa.eu/regional_policy/sources/docoffic/official/regulation/pdf/2007/general/ce_1083(2006)_en.pdf

Guidelines:

http://ec.europa.eu/regional_policy/sources/docoffic/working/doc/exante_sf2007-2013_en.pdf

http://ec.europa.eu/regional_policy/sources/docoffic/working/doc/indic_sf2007-2013_en.pdf

Presentations:

http://www.degeval.de/calimero/tools/proxy.php?id=13203#search=%22evaluation%20during%202007-2013%22

Bibliography

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

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. From a democratic perspective, accountability is an important dimension of evaluation. Public authorities are progressively increasing their requirements for transparency vis--vis tax payers, as to the sound use of funds they manage. In this spirit, evaluation should help to explain where public money was spent, what effects it produced and how the spending was justified. Those benefiting from this type of evaluation are political authorities and ultimately citizens. For example, a training organisation reports on the number of trainees who benefited from its services and the qualifications obtained. A managing authority reports on the cost per job created due to the intervention. The European Commission publishes a report on progress made in terms of economic and social cohesion. Citizens have access to the report.

Accountable Body

The organisation responsible in a contractual sense for the intervention, with the key responsibilities of ensuring that the programme is managed in accordance with required standards of financial probity, and in line with the agreed action plan, and any programme specific funding criteria.

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

Additionality has been one of the principles of EU Structural Fund interventions. Additionality means that Community support for economic and social development is not substituted for 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. This term additionality is also used to describe the net effects of an intervention identified in an evaluation.

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 general criteria and standards, known and clarified beforehand. For example, in the case of assistance to a SME, an audit will check whether eligibility criteria have been met and whether the beneficiary firms have complied with the rules governing the use of assistance. The main purpose of an audit is to ensure compliance. The idea is to obtain a dissuasive effect. 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. Others assess whether the results and performance are similar to other comparable interventions. In the latter case, there can be a closer similarity between this variant of auditing and evaluation.

Baseline

State of the economic, social or environmental context, at a given time (generally at the beginning of the intervention), and against which changes will be measured. The basic situation is described by context indicators which describe the economy, socio-economic environment, concerned groups, etc.

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. Beneficiaries receive support, services and information, and use facilities created with the support of the intervention (e.g. a family which uses a telephone network that has been improved with public intervention support, or a firm which has received assistance or advice). 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. The term "addressee" was used in the MEANS collection because it was felt the term "beneficiary" is associated with the idea of a positive effect, which can be misleading because evaluation may identify negative impacts when investigating effects on direct 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 involve providing details of the proposed research methodology.

Case Study

In-depth study of data on a specific case (e.g. a project, beneficiary, town). The case study is a detailed description of a case in its context. It is an appropriate tool for the inductive analysis of impacts and particularly of innovative interventions for which there is no prior explanatory theory. Case study results are usually presented in a narrative form. A series of case studies can be carried out concurrently, in a comparative and potentially cumulative way. A series of case studies may contribute to causal and explanatory analysis.

Causal analysis

The study of relations of cause and effect which link a public intervention to its impacts. Causality analysis may be inductive. In this case, it investigates the mechanisms likely to produce impacts, as well as confounding factors likely to have an influence. Causality analysis may also be deductive (or hypothetico-deductive). In this case, it examines whether assumptions about impacts are not contradicted by the facts. It may also supply a quantitative estimation of impacts.

Citizen's juries

A panel of 12-16 members of the public brought together to consider a public policy issue posed by an authority or evaluation commissioning body. The jury might sit over 3/4 days and be assisted by neutral facilitators - and are expected to reach a conclusion or form a judgement.

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). Secondly, comparability can be extended to key indicators, that is, indicators which can be used in several regions and sectors of intervention.

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. Complementarity may be functional, if the objective is functional (e.g. developing a sector or creating a network). It may be territorial, if the objective concerns a territory (e.g. integrated local development policy).

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. The completion rate of an intervention is the weighted average of the completion rates of the projects comprising that intervention. The completion rate is a core indicator for monitoring the implementation of an intervention.

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

Factors independent of a public intervention which are partly or entirely the cause of changes observed among beneficiaries (gross effects). A confounding factor can hinder or amplify an expected impact. For example, employment increased by 10% in a group of assisted firms. However, this increase could not be imputed entirely to the public intervention because the assisted firms also benefited from a favourable macro-economic context (confounding factor). In another example, close to 80% of the trainees found a job after one year. However, this high placement rate was partly imputable to the initial qualifications of the trainees which were very high at the time of their recruitment (skimming-off effect). When a causality analysis is carried out as part of an evaluation, it distinguishes between effects imputable to the public intervention (net effects) and those imputable to confounding factors.

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 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 human knowledge is not developed exogenously but through individual experiences and social interactions. It claims that what observation allows us to describe are in part at least constructed by the observer - the world is not fully observable. 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 as much as with studies and externally generated data.

Context

The socio-economic environment in which an intervention is implemented. The term is used in its broadest sense. For example, in the case of interventions in favour of SMEs, the context includes the macro-economic situation and the framework conditions (tax laws, company law, etc.).

Control group

Comparison group consisting of eligible people or organisations which have been excluded from all 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. It has also been exposed to the same variations in the socio-economic context (confounding factors). When a group of participants and a control group are compared, the influence of confounding factors is the same on both sides (provided the two groups are large enough).

Correlation

The fact that two variables systematically evolve in the same direction or in opposite directions. If there is little covariance, there is no statistically reliable relationship between them. 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 as it would have been in the absence of a public intervention. The difference between counterfactual and actual outcome is the impact of policy. By its very nature, a counterfactual is always an estimation. The simplest method is to use the initial situation ("baseline") as the counterfactual. For example, 100 SMEs receive investment support, between them they increase their capital stock from EUR20 million to EUR30 million. In this simple scenario, EUR20m is the baseline and EUR30m-EUR20m = Eur10 million is the estimated impact of assistance. Although the baseline is the simplest method, where possible other tools are preferable. For example, it is found that a control group of 100 similar SMEs from the same region increased their capital stock from EUR20 million to EUR27 million in the same period without assistance. In this case, the counterfactual is EUR27 million and the estimated impact EUR3 million. Although control groups are said to be the gold standard for establishing a counterfactual, practicality often dictates the use of other tools. Common examples are shift-share analysis and simulation using econometric models.

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 in a comparative way, at a given time, on several categories of individuals or facts. For example: comparative unemployment rates in European regions; comparative placement rates for women and men.

Deadweight

Change observed among direct beneficiaries following the public intervention, or reported by direct addressees as a consequence of the public intervention, that would have occurred, even without the

intervention. For example: a farmer received assistance for the building of a self-catering cottage. In the survey he stated that the support had enabled him to create better quality facilities, but that he would have built the cottage, even without support. Thus, there is deadweight since the the construction of the cottage cannot be imputed entirely to the intervention. The estimation of deadweight usually necessitates a survey of direct beneficiaries, preferably with a comparative analysis of non-participants.

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.

Difference in differences

Difference in differences is a relatively simple way of getting much of the rigour of experimental methods. A control group is selected which shares some key characteristics with the intervention group. The two groups are 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. Difference in difference therefore avoids some of the practical difficulties of experimental methods – notably having to deny assistance to the controls or having to set up a randomised control group at the beginning of the programme. A disadvantage is that although control and intervention groups are matched on some key characteristics, the difference in performance may be driven by differences in unobserved (or unobservable) characteristics. Where possible, other methods (e.g. propensity score matching) should be used as a check.

Direct effect

Effects of a public intervention on its direct beneficiary, excluding all repercussions on other groups. For example: investment support has direct effects on assisted businesses (production costs and capacity; creation or maintenance of jobs). New infrastructure has a direct effect on the people and enterprises which use it.

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. Displacement effects may be intended (e.g. displacement of a public administration from the capital to a 'lagging' region) or unintended (e.g. 10% of the jobs created by a regional development programme resulted in the disappearance of jobs in other eligible regions; a firm used programme assistance to move its premises from the centre to the outskirts of a town). When they are not intended, displacement effects must be subtracted from gross effects to obtain net effects. The term is sometimes used to refer to the effects of an intervention affecting one beneficiary being at the expense of another within the same territory.

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.

Econometric analysis

The application of econometric models used to simulate the main mechanisms of a regional, national or international economic system. A large number of models exist, based on widely diverse macro-economic theories. This type of tool is often used to simulate future trends, but it may also serve as a tool in the evaluation socio-economic programmes. In this case, it is used to simulate a counterfactual situation, and thus to quantitatively evaluate net effects on most of the macro-economic variables influenced by public actions, i.e.: growth, employment, investment, savings, etc. The models are generally capable of estimating demand-side effects more easily than supply-side effects. Econometric analysis is also used in the evaluation of labour market interventions.

Economic and social cohesion

This has two meanings. 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). More generally social cohesion implies a level of social integration and solidarity that holds together societies and communities.

Economic Impact Assessment

Economic impact assessment is about tracking or anticipating the economic impact of an intervention. It depends on analysing the cause and effect of an intervention and is important in project appraisal. It can be undertaken before, during or after projects to assess the amount of value added by a given intervention and whether it is justified.

Effect

Socio-economic change resulting directly or indirectly from an implemented intervention. Effects include the results and impacts of an intervention, whether positive or negative, expected or not. The term should not be used to describe outputs.

Effectiveness

The fact that expected results or effects have been obtained and that objectives have been achieved. Effectiveness can be assessed by answering the following questions, for example: "Could more effects have been obtained by organising the implementation differently?" or "Which are the most successful operators or measures?". An effectiveness indicator is calculated by relating an output, result or impact indicator to a quantified objective. For example: 85% of the the objective for the number of firms created was achieved; the placement rate obtained by operator A is better than that obtained by operator B. For the sake of clarity, it may be useful to specify whether one is referring to the effectiveness of outputs, results or impacts.

Efficiency

Effects obtained at a reasonable cost. Efficiency may be assessed by answering the following questions, for example: "Could more effects have been obtained with the same budget?" or "Have other interventions obtained the same effects at a lower cost?". An indicator of efficiency is calculated by dividing the budgetary inputs mobilised by the quantity of effects obtained. For example: the average cost of training a person who has been jobless for a long time is 2,000 euro; the intervention should achieve a cost per job created of less than 30,000 euro. For the sake of clarity, it could be useful to specify whether the efficiency referred to relates to outputs, results or impacts. The efficiency of outputs may be measured in terms of the unit cost.

Related Terms: Cost Effectiveness, Unit-cost, Value for Money

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. For example, is a programme targeted on the unemployed reaching that population? European regions eligible for support attributed to lagging development areas must have a GDP lower than 75% of the European Union average. 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, so that the individual concerned can keep his or her job or find a (new) job in reasonable conditions in a reasonable length of time. A public intervention in favour of employability concerns human resource development and particularly training. 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).

Employment effect

Improvement in employability, creation and maintenance of jobs, or structural modification of the labour market, following an intervention.

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 compulsory step in certain countries 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. This impact relates to the principle of equal rights and equal treatment of women and men. It means, first, that everybody is free to develop their personal aptitudes and to make choices without being limited by stereotyped gender roles and, secondly, that particular differences in behaviour, aspirations and needs, between women and men, are not to be valued too highly nor considered too critically. The principle of equal opportunities may require unequal treatment to compensate for discrimination. The evaluation of impacts on equal opportunities includes the mainstreaming of gender at all stages. Equal opportunities can be applied to characteristics other than gender, race, ethnicity, sexuality etc.

Evaluability assessment

Technical part of the 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. It is closely linked with examinations of programme theory and programme logic insofar as evaluability depends on the coherence of the programme's logic and the plausibility of its interventions and implementation chains.

Evaluation

Judgement on the value of a (usually) public intervention with reference to criteria and explicit standards (e.g. its relevance, efficiency, sustainability, equity etc.). The judgement usually concerns the needs which have to be met by the intervention, and the effects produced by it. The evaluation is based on information which is specially collected and interpreted to support the the judgement. For example: evaluation of the effectiveness of a programme, cost-benefit evaluation of a project, evaluation of the validity of a policy, and evaluation of the quality of a service delivered to the public. Certain definitions of evaluation exclude the judgement dimension and limit evaluation to an assessment, description or measurement of the intervention's effects. Other, more restrictive

definitions, limit evaluation to the ex post estimation of effects. In certain contexts, evaluation focuses not on a public intervention but on a public organisation (e.g. evaluation of a university, hospital or EU institution). More generally, the term is used in human resource management for the evaluation of a person (e.g. annual evaluation interview) or in the financial domain to estimate the value of an enterprise. Journalistic use of the word is often in the sense of a vague estimation (the number of victims is estimated at 1,000 people) or a monetary estimation (this table is valued at 1 million euros). Many definitions of evaluation start with different purposes (management, institutional strengthening, accountability, learning) and link this to the uses of evaluation outputs.

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. These are usually posed by those commissioning an evaluation thought this will be on behalf of key stakeholders such as programme managers, policy makers, parliaments and citizens. Evaluation questions normally feature in the terms of reference which the evaluation team will have to answer. Evaluation questions have three dimensions: descriptive (what happened?), causal (to what extent is what has happened really an effect of the intervention?) and normative (is the effect satisfactory?). An evaluation generally has several questions. Sometimes when evaluation questions are not clearly stated or when the programme logic is uncertain and needs to be clarified, evaluation questions are restated in an inception report to ensure they are well focused.

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 a vis the commissioning body or programme managers. Evaluation teams may bring together a group of evaluators drawn from a single or several organisations (consortium). The evaluator can also be an individual or even in some formulations an evaluation authority that makes the final judgement on the basis of the work produced by an evaluation team.

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. Ex ante evaluation mainly concerns an analysis of context, though it will also provide an opportunity for

specifying the intervention mechanisms in terms of what already exists. 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, etc. Moreover, it provides the necessary basis for monitoring and future evaluations by ensuring that there are explicit and, where possible, quantified objectives.

Ex post evaluation

Evaluation which recapitulates and judges an intervention when it is over. It aims at accounting 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 results and impacts. It also tries to draw conclusions which can be generalised to other interventions. For impacts to have the time to materialise, ex post evaluation needs to be performed some time after implementation. In order to assess impacts, ex post evaluations are likely to involve field studies that track change over time as well as snap-shots such that one time surveys will provide.

Expert panel

Work group which is specially formed for the purposes of the evaluation and which may meet several times. The experts are recognised independent specialists in the evaluated field of intervention. They may collectively pronounce a judgement on the value of the public intervention and its effects. An expert panel serves to rapidly and inexpensively formulate a synthetic judgement which integrates the main information available on the programme, as well as information from other experiences.

Explanatory theory

All the assumptions likely to explain changes observed following the public intervention (gross effects). 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 between outputs, results and impacts. It also covers any other causes likely to explain gross effects, 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. Inductive analysis techniques also contribute towards the construction or improvement of explanatory theory.

External coherence

Correspondence between the objectives of an intervention and those of other public interventions which interact with it. If a national policy and a 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

Evaluation of a public intervention by people not belonging to the administration responsible for its implementation. For example, a team composed of private consultants, researchers or people belonging to public organisations unrelated to those responsible for the intervention. External evaluations are often seen as more impartial and independent and therefore as more credible. It is possible in some circumstances for internal evaluators supported by a culture of evaluation to be similarly credible. There are also constraints on the access and internal credibility of external evaluators which have to be balanced against their advantages.

External validity

Quality of an evaluation method which makes it possible to obtain conclusions that can be generalised to contexts (groups, areas, periods, etc.) other than that of the intervention being evaluated. For example, evaluation makes it possible to conclude (1) that in directly or indirectly assisted SMEs the intervention helped to double the innovation rate and (2) that similar assistance, attributed to SMEs in other regions, would probably have exactly the same effect. Only strong external validity allows one to extrapolate from lessons learned during the implementation of the evaluated intervention. It is particularly sought after when the evaluation aims at identifying and validating best practice. External validity is also required when the evaluation uses conclusions of evaluations performed on similar interventions.

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.

Extrapolation coefficient

Ratio used to estimate an impact through the use of monitoring data. For example, if the amount of investments by assisted SMEs is known, an extrapolation coefficient can be used to estimate the annual value added generated. In this case, the extrapolation coefficient may be the ratio of annual value added per euro invested. This ratio will be drawn from relevant regional or sectoral statistics. If the evaluator uses existing extrapolation coefficients, it is important for it to explain and justify its choices. A coefficient can be deduced from a questionnaire survey (provided the sample is representative) conducted as part of the same evaluation or a preceding one. A coefficient may also be taken from an evaluation or research on similar interventions, provided the conclusions can be generalised (external validity). A coefficient can be established for the needs of an evaluation by an expert panel or by a network of experts, using the Delphi panel technique.

Factor analysis

Statistical analysis tool used to identify all correlations within a large quantity of data concerning many individuals (e.g. survey results), and to represent them in a simplified way in the form of a two- or three-dimensional space. The tool reveals groupings and suggests relations of cause and effect. It is an inductive causality analysis tool.

Feedback

Feedback is a process by which evaluation formulations, 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. For example, the same programme can predict outputs in the field of research and development, in that of transport infrastructure, in that of training, and so on. Within the framework of European cohesion policy, fields of intervention have been grouped together into three main categories: basic infrastructure, productive environment and human resources. The term policy area has been used to reflect the wider range of potential interventions linked to socio economic development that may be applied in the context of enlargement of the EU.

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 a 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

Change observed following a public intervention, or an effect reported by the direct beneficiaries. A gross effect appears to be the consequence of an intervention but 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 claimed to have created 500 jobs owing to the support (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).

Impact

A consequence affecting direct beneficiaries following the end of their participation in an intervention or after the completion of public facilities, or else an indirect consequence affecting other beneficiaries who may be winners or losers. Certain impacts (specific impacts) can be observed among direct beneficiaries after a few months and others only in the longer term (e.g. the monitoring of assisted firms). In the field of development support, these longer term impacts are usually referred to as sustainable results. Some impacts appear indirectly (e.g. turnover generated for the suppliers of assisted firms). Others can be observed at the macro-economic or macro-social level (e.g. improvement of the image of the assisted region); these are global impacts. Evaluation is frequently used to examine one or more intermediate impacts, between specific and global impacts. Impacts may be positive or negative, expected or unexpected.

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, in order to give an opinion on the "Res publica". Impartiality is an essential element of the quality of an evaluation.

Implementation

The operational process needed to produce expected outputs. In the context of EU socio-economic programmes, implementation comprises all or part of the following tasks: mobilising and distributing allocated inputs; assigning management responsibilities to operators; selecting calls for tenders for project promoters; and, lastly, selecting and financing projects. To monitor and improve implementation, a monitoring committee is set up, a system of information monitoring is launched, and audits and evaluations are performed.

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.

In house evaluation

Evaluation by an evaluation team belonging to the administration responsible for the programme. Internal evaluation may be more independent if the evaluation team has no hierarchical relation with the actors implementing the intervention. Otherwise, it would be a form of self-evaluation.

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

Measurement of: an objective to achieve; a resource mobilised; an output accomplished; an effect obtained; or a context variable (economic, social or environmental). The information provided by an indicator is a quantitative datum used to measure facts or opinions (e.g. percentage of regional enterprises which have been assisted by public intervention; percentage of trainees who claim to be satisfied or highly satisfied). An indicator must, among other things, produce simple information which is communicable and easily understood by both the provider and the user of the information. It must help the managers of public intervention to communicate, negotiate and decide. For that purpose, it should preferably be linked to a criterion on the success of the intervention. It should reflect as precisely as possible whatever it is meant to measure (validity of construction). The indicator and its measurement unit must be sensitive, that is to say, the quantity measured must vary significantly when a change occurs in the variable to be measured. Indicators may be specially constructed by the evaluation team and quantified by means of surveys or statistical data. They are often borrowed from the monitoring system or statistical series. An indicator may be elementary or derived from several other indicators in the form of ratios or indices.

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 effects are produced in the short or medium term, positively or negatively. The *raison d'être* of an intervention is to produce positive change for all its direct and indirect beneficiaries.

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 includes 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. For example, sixty people worked on implementing the programme; 3% of the project costs were spent on reducing effects on the environment. Monitoring and evaluation focus primarily on the inputs allocated by public authorities and used by operators to obtain outputs. Private inputs mobilised by assisted firms, for example, are considered to be results of public intervention. The above definition 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. The term "financial outputs" is sometimes used in the sense of consumption of budgetary inputs.

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. The tool can be used when a table of inputs and outputs is available. This is usually the case at the national level but more rarely so at the regional level. The tool is capable of estimating demand-side effects but not supply-side effects.

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

Quality of an evaluation method which, as far as possible, limits biases imputable to data collection or processing techniques. For example, an intervention which directly or indirectly assisted SMEs, led to the doubling of the innovation rate per year for 1,000 jobs. This type of conclusion is sounder if the collection and analysis of data have very precisely taken into account all the specific aspects of the intervention and its context: categories of assisted firms, forms of innovation specific to the region, etc. To obtain better internal validity, it is necessary to strictly control a series of parameters, otherwise an artificial situation may be created which would limit the possibility of generalising the conclusions (less external validity).

Interval data

Interval data is continuous data where differences are interpretable, but where there is no "natural" zero. One example that can be applied here is temperature in Fahrenheit degrees.

Intervention

Any action or operation carried out by public authorities regardless of its nature (policy, programme, measure or project). Interventions generally take one of the following forms: construction or improvement of facilities, subsidies, loans, secured loans, consulting or provision of services, training, information campaigns, setting up of rules and regulations. In this Volume, 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 direct 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 additional effects (see net effect). Nor do they refer to the principle of additionality which applies to European social and economic cohesion policy (see additionality).

Logic models

Generic term that describes various representations of programmes linking their contexts, assumptions, inputs, intervention logics, implementation chains and outcomes 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 - ToC - 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 outputs, results, and specific and global impacts. Each level of objective is associated with one or more verifiable indicators of success, as well as with the conditions and risks influencing success or failure (confounding factors). This tool is frequently used for the evaluation of development support projects. By contrast, it is ill-suited to the evaluation of complex socio-economic programmes because its level of simplification is high.

Longitudinal data

Time series relating to repeated comparative observations of the same individuals (or other beneficiary), at regular intervals, during a given period. In the evaluation context, these data are obtained by observing the evolution of a sample of beneficiaries.

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. For example, equal opportunities is mainstreamed in evaluation in so far as evaluation systematically distinguishes men and women among the direct and indirect beneficiaries of a public intervention, and among operators and decision-makers, when this distinction concerns needs, results and impacts. Mainstreaming may also apply to other impacts considered to be priorities. The term is also used to refer to the process of extending innovative practices to mainstream programmes and policies.

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. The political dimension of the evaluation act is overlooked. The quality of managerial evaluation is based on two principles: the independence of the evaluator and the objectivity of the method. For example, during the 1980s, most functions of the British administration were privatised or delegated to semi-independent agencies. Demands for effective services grew at the same time. This led to the development of Value For Money evaluations.

Matching pairs

Technique for constructing a comparison group. For each individual in the sample of direct beneficiaries being surveyed, an individual is chosen who resembles it as closely as possible and who has not participated in the intervention.

Matrix of cross impacts

Tool used to highlight potential complementarity, duplication and conflict between the impacts of the different components of an intervention, or between the impacts of different interventions. For example, the tool presents the different measures of a programme in lines and columns. In each cell a score between 3 and +3 is given to describe the potential synergy between a measure situated on a line and a measure situated in a column.

Measure

Within the framework of European economic and social cohesion policy, the basic unit of programme management, consisting of a set of similar projects and disposing of a precisely defined budget. Each measure has a particular management apparatus. Measures generally consist of projects. Many measures are implemented through a process of Calls for Proposals and subsequent appraisal.

Measurement unit

Used to observe a phenomenon, change or variable, and to place it on a quantitative scale. A measurement unit allows for quantification. An elementary indicator is associated with a measurement unit and has only one dimension (e.g. km of motorway; number of training courses).

Some measurement units are divisible and others not (e.g. 20.3km were built; 30 trainees were qualified). Measurement units must be harmonised if indicators are to be comparable.

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 results and findings. Meta evaluations can focus on programme 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 generalisation 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).

Micro economics

The study of the behaviour of small economic units, such as individual consumers, households or firms with respect to consumption, production and other socio-economic decisions.

Mid-term evaluation

Evaluation which is performed towards the middle of the period of implementation of the intervention. This evaluation critically considers the first outputs and results, which enables it to assess the quality of the monitoring and implementation. It shows the translation into operational terms of initial intentions and, where relevant, points out de facto amendments to objectives. Through comparison with the initial situation, it shows the evolution of the general economic and social context, and judges whether the objectives remain relevant. It examines whether the evolution of policies and priorities of other public authorities raises problems of coherence. It also helps to prepare adjustments and reprogramming, and to argue them in a transparent manner. Mid-term evaluation not only relies strongly on information derived from the monitoring system, but also on information relating to the context and its evolution. 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 intermediate evaluation. Other intermediate evaluations may be performed during the first or last years of implementation.

Monitoring

An exhaustive and regular examination of the resources, outputs and results of public interventions. Monitoring is based on a system of coherent information including reports, reviews, balance sheets, indicators, etc. Monitoring system information is obtained primarily from operators and is used essentially for steering public interventions. When monitoring includes a judgement, this judgement refers to the achievement of operational objectives. Monitoring is also intended to produce feedback and direct learning. It is generally the responsibility of the actors charged with implementation of an intervention. For example: monitoring of the consumption of budgets, monitoring of the meeting of deadlines, monitoring of the percentage of SMEs in beneficiary firms, monitoring of the level of qualifications obtained by trainees.

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. Mid-term evaluation may involve a survey of beneficiaries, to reveal their needs and opinions. 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 apparent changes or gross effects. To evaluate net effects, based on gross effects, 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 since they are produced by confounding factors. For example, the number of employees in assisted firms appears to be stable (change or gross effect equal to zero). However, it is estimated that without support there would have been 400 redundancies. Thus, 400 jobs were maintained (net effect).

Network effect

Secondary effect which spreads through special and lasting relations between businesses in the same territory or branch, and between those same businesses and public or para-public organisations such as research centres, universities, etc. Network effects are partially based on non-market interpersonal relations. Thus, they are in part "externalities".

Nominal data

The lowest level of data measurement. Categorical data where the order of the categories is arbitrary. For example, race/ethnicity has values 1=White, 2=Hispanic, 3=American Indian, 4=Black, 5=Other.

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

Clear, explicit and initial statement on the effects to be achieved by a public intervention. If the objectives are not stated explicitly an evaluation (and particularly ex ante evaluation) may help to clarify them. A quantified objective is stated in the form of indicators and a qualitative objective in the form of descriptors, e.g. 30% of all outputs must be accomplished by the end of the third year; the public intervention must first benefit the long-term unemployed. Specific objectives concern the results and impacts of an intervention on direct beneficiaries. A global objective corresponds to the aim of the intervention. The aim of an intervention is to produce an impact expressed in global terms, e.g. reducing regional disparities in development levels. Objectives may also be intermediate. Objectives which specify outputs to be produced are called operational objectives. If the objectives of a public intervention have not been clearly defined beforehand, the evaluation can try to clarify them afterwards. In that case, it is preferable to refer to implicit objectives. Objectives may incorporate targets.

Objective tree

Hierarchical classification of the objectives of a public intervention, linking each specific objective to its global objective. The objectives tree of a programme makes the overall logic explicit. It is used for clarification.

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. For example, an on-going evaluation of development support for tourism has successively considered the following questions: "how has the public reacted to the proposed support?" (first year); "is the evolution of the tourist market making the assistance useless?" (third year); "which of the assisted projects can be considered cases of best practice to reproduce?" (fourth year). 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.

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, with audio-visual recordings where relevant, on a few typical situations of implementation of the intervention. 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. In socio-economic development participatory approaches are especially relevant because they support autonomy and self confidence rather than encourage dependency.

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 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 fact that effects were obtained at a reasonable cost and that the beneficiary are satisfied with them. For example: a water purification programme has a high performance rate if the per capita cost is limited compared to similar interventions, if the purification plants built comply with quality standards, and if the rivers concerned are less polluted. Efficiency and performance are two similar notions, but the latter extends, more broadly, to include qualitative dimensions. In certain contexts, performance concerns outputs and results but not impacts. In other contexts, the term applies mainly to either outputs (World Bank Operations Evaluation Department) or impacts. The meaning of the word performance is not yet stable; it is therefore preferable to define it whenever it is used.

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 4% of appropriations allocated to each Member State for the 2000-2006 period were placed in reserve until 2003. They were then distributed

to the best-performing programmes. The measure was designed to motivate the Fund recipients. Each Member State made proposals to the Commission on the basis of monitoring indicators that it has itself introduced.

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. In France, ad hoc evaluation authorities are frequently set up to manage evaluation work. For example, in the framework of national evaluation of the prevention of natural hazards, the evaluation authority was composed of actors from diverse origins: ministries, local authorities, public agencies, associations, etc. Within the framework of policies co-financed by the State and the regions, the steering of certain evaluations is entrusted to political and administrative officials at both levels, in partnership with representatives of the groups concerned.

Policy

A set of different activities (programmes, strategies, procedures, laws, rules) directed towards a single goal or general objective. These activities are often accumulated incrementally through the years. European economic and social cohesion policy is scheduled according to a precise time-frame, with a multi-annual budget. This is not the case for the majority of policies, in the traditional sense of the term.

Policy area

The term used to describe a specific policy domain of interventions, laws and regulations, programmes, projects etc. i.e. economic development, social policy.

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.

Policy priority

The funding authorities' wish that evaluation should examine certain impacts which were not stated as objectives when the intervention was launched, but which represent political priorities at that level. For example, in the framework of its economic and social cohesion policy, the European Union demands that evaluations systematically take into account impacts on the environment, on the competitiveness of SMEs, on the creation and maintenance of jobs, and on equal opportunities between men and women.

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). Primary data are particularly useful for analysing impacts and needs in depth.

Priority Evaluation

The priority-evaluator technique was developed as a way of involving the public in decisions about complicated planning issues. The method is an attempt to combine economic theories with survey techniques in order to value unpriced commodities, such as development or environmental conservation. It is used to identify priorities in situations where there is likely to be a conflict of interest between different people or interest groups, and the choice of any option will require a trade-off. The priority evaluator technique is designed around the identification of a set of options comprising varying levels of a given set of attributes. The basis of the technique is to let the respondent devise an optimum package, given a set of constraints. The method allows the research to identify the cost of moving from one level of each attribute to another, and the respondent is invited to choose the best package, given a fixed budget to spend. The analysis is based on neo-classical microeconomic assumptions about consumer behaviour (e.g. the equation of marginal utility for all goods), thus arriving at respondents ideally balanced preferences, constrained financially, but not limited by the imperfections and limitations of the market place.

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. A programme is always under the responsibility of an authority or several authorities who share the decision-making. Within the framework of European social and economic cohesion, programmes are generally broken down into axes, measures and projects.

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

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 economic and social 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 tends to be very complicated but at its heart are 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. 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. A simple practical hurdle is that it depends on having good background information. A more hidden – but very real – danger is that the complexity of the matching process may generate spurious certainty. Comparison with other methods should be used to check the robustness of results.

Qualitative indicator

A description, in the form of a concise, clear and stable statement, of an objective to achieve or an impact obtained. For example, one can examine the impact of a measure in terms of equal opportunities, and conclude by choosing a pre-established descriptor such as: (1) "the measurement enabled most women addressees to qualify for jobs considered to be reserved for men", or (2) "the measure attracted a large proportion of women addressees who wanted to qualify for jobs considered to be reserved for men". The organisation of descriptors in the form of a structured grid may constitute the first step in the construction of an indicator. If several descriptors have been established beforehand, they can be used to construct an observation grid. By means of this grid a phenomenon or change can be observed and described in a qualitative and structured way. Evaluation cannot afford to exclude from its scope of analysis an important objective or impact simply because it is difficult to measure quantitatively when in fact it is considered to be important. In that case, it is preferable to collect qualitative data and to structure them by means of descriptors.

Quantitative indicator

Facts and figures, describing things using numbers. Tells you what happened, e.g. number of beneficiaries, % who are female.

Questionnaire survey

A set of structured questions addressed to a sample of people or, exhaustively, to a group of people. A questionnaire consists of a list of questions of which the sequence is determined in advance. Depending on whether the questions are open or closed, the interviewee is either free to formulate his or her answers as he or she wishes, or is given predefined statements (descriptors) from which to choose. A questionnaire can be undertaken by post, telephone, e-mail or face-to-face interview. If the questionnaire survey is exhaustive or if it concerns a representative sample, and provided that most of the questions are closed, it may be used for gathering quantitative data and for informing indicators.

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. Certain specialized statistics, such as a geometric mean and a coefficient of variation can only be applied to ratio data.

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

Statistical tool used to make a quantitative estimation of the influence of several explanatory variables (public intervention and confounding factors) on an explained variable (an impact). Regression analysis is a tool for analysing deductive causality. It is based on an explanatory logical model and on a series of preliminary observations. The tool can be used in varying ways, depending on whether the variables of the model are continuous or discrete and on whether their relations are linear or not.

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. Sound reliability implies exhaustive data collection and the appropriateness of the evaluative questions asked. This notion is important not only for primary data but also for secondary data, the reliability of which must be carefully checked.

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.

Results

Advantage (or disadvantage) which direct beneficiaries obtain at the end of their participation in a public intervention or as soon as a public facility has been completed. Results can be observed when an operator completes an action and accounts for the way in which allocated funds were spent and managed. At this point s/he may show, for example, that accessibility has been improved due to the construction of a road, or that the firms which have received advice claim to be satisfied. The operators may regularly monitor results. They have to adapt the implementation of the intervention according to the results obtained.

Sample

Part of a group selected for a survey. In its statistical sense, the word "population" applies to all the people, projects or objects likely to be observed, e.g. all the assisted firms or individuals; all the firms in a region; or an entire targeted group. If the sample is representative, the survey results can be generalised to the entire surveyed population (internal validity) or to similar populations (external validity). A sample is representative if it has the same distribution of characteristics as the population from which it is taken and if it is large enough (generally at least a thousand individuals). Increasing sample size improves the statistical significance of survey findings.

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.). In the context of European cohesion policy, the evaluation object may be a policy, programme, a measure or a project. The term public intervention is applied, generically, to any evaluation object.

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. Self evaluation is especially suited to assist managers, promoters and immediate stakeholders improve their own performance and steer their programmes in the course of implementation.

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 Cost Benefit Analysis

Cost-benefit analysis is well established and widely used method for evaluating project proposals. Some argue that the difficulty with the traditional cost-benefit measure is that it treats all income effects equally, regardless of which socio-economic group it accrues to. Thus, projects / policies can also be subjected to a Social Cost-Benefit Analysis in order to determine their impact on different socio-economic groups in addition to their overall impact on a geographical area. There are key differences in conducting a social cost benefit analysis, as opposed to a standard cost-benefit analysis. All relevant costs and benefits need to be identified and quantified. In addition, a weighting system needs to be developed to recognise the impact on different groups and the analysis needs to arrive at an appropriate discount rate.

Social partners

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

Social survey

Surveys are used to collect a broad range of information (quantitative and qualitative) about a population. The emphasis is usually on quantitative data.

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

Said of a sample size which is large enough to be considered representative of the overall population being studied. Significance tests are performed to see if the null hypothesis can be rejected. If the null hypothesis is rejected, then the effect found in a sample is said to be statistically significant. If the null hypothesis is not rejected, then the effect is not significant. The experimenter chooses a significance level before conducting the statistical analysis. A statistically significant effect is not necessarily practically significant.

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. An evaluation steering committee may notionally involve any person who is potentially a user of its recommendations, any person who has an interest in the information produced, and any person who is likely to win or lose in the course of the programme. The main categories of stakeholders are funding authorities, managers, operators and concerned groups. Stakeholders invited to join an evaluation steering committee improve the relevance of the questions asked and their presence makes the evaluation more credible. A danger in steering committees is that they might interfere in the evaluation and seek to influence conclusions rather than ensure its accurate use of information, an understanding of a particular context and that the evaluation team is true to its terms of reference.

Strategic Environmental Assessment

A similar technique to Environmental Impact Assessment but normally applied to policies, plans, programmes and groups of projects. Strategic Environmental Assessment provides the potential opportunity to avoid the preparation and implementation of inappropriate plants, programmes and projects and assists in the identification and evaluation of project alternatives and identification of cumulative effects. Strategic Environmental Assessment comprises two main types: sectoral strategic environmental assessment (applied when many new projects fall within one sector) and regional SEA (applied when broad economic development is planned within one region).

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 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 EUs competitive position and, consequently, to improve the prospects of its citizens. The total budget for the Structural Funds amounts to 195 billion Euro in 2000-06, divided between the European Regional Development Fund (ERDF), the European Social Fund (ESF), the European Agricultural Guidance and Guarantee Fund (EAGGF) and the Financial instrument for Fisheries

Guidance (FIFG). Most Structural Fund assistance is granted in the form of non-repayable grants or direct aid, and to a lesser degree refundable aid, interest-rate subsidies, guarantees, equity participation, and participation in venture capital.

Structural Indicators

At the end of October 2001, the European Commission adopted a communication on Structural Indicators. The aim of these indicators is to measure the EUs progress towards becoming the most competitive and innovative economy in the world, in order to increase its growth potential and achieve full employment. The indicators fall into five categories: general economic context, employment, innovation, economic reforms, social cohesion, the environment and sustainable development.

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

That which justifies public authority's decision to implement an intervention rather than to leave it up to private initiative or another public authority. The principle of subsidiarity justifies public intervention when there are shortcomings in the private sector and when other public administration levels would not have been as effective. Subsidiarity is part of the rationale of an intervention. In the European context, subsidiarity means, for example, that the Community acts in those cases where an objective can be achieved better at the European level than at the level of Member States taken alone. This corresponds to the notion of Community value added which is frequently considered a criterion to take into account during the evaluation of EU funded programmes.

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 positive. An evaluation determines, with regard to the objectives of the intervention, whether the substitution effect can be considered beneficial or not. When it is not beneficial, the substitution effect must be subtracted from gross effects.

Summative Evaluation

Evaluation performed after a programme has been implemented to assess its results and impacts, whether it achieved its aims and objectives and whether it was beneficial overall to those it was intended to benefit.

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.

Sustainability

The ability of effects to last in the middle or long term. Effects are sustainable if they last after the funding granted by the intervention has ceased. They are not sustainable if an activity is unable to generate its own resources, or if it is accompanied by negative effects, particularly on the environment, and if that leads to blockages or rejection.

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

All the people and organisations for which an intervention is directly intended (e.g. people unemployed for over a year; entrepreneurs). An eligible public is composed of all potential direct beneficiaries. Only a part of them are affected by the intervention (see the notion of coverage rate). This concept is more limited than the term concerned public which also includes foreseen or unforeseen beneficiaries.

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 is going to produce its impacts 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. Overall variations in the characteristics of a given population are observed over time. 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.

Unintended

An impact which is revealed by evaluation but was not foreseen when an intervention was launched. Unexpected effects are revealed by inductive analysis techniques (particularly case studies). They are not part of the objectives. A positive unexpected effect may become an implicit objective. Negative unexpected effects are also called perverse effects. For example, if assistance is granted for the development of a territory and, for that purpose, the territory is classified "disadvantaged", it will become less attractive for business as a result.

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). See also scope of evaluation, above.

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 squared deviation of each number from its 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. The weighting of criteria can be formalised by expressing it as a percentage (the total being 100%). Multicriteria analysis also uses weighting.

Welfare

Welfare can be either be 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.