

**THE EVALUATION OF
SOCIO-ECONOMIC DEVELOPMENT**

The GUIDE

December 2003

Tavistock Institute

in association with:

GHK

IRS

CONTENTS

INTRODUCTION	1
PART 1 THE CONTRIBUTION OF EVALUATION TO SOCIO-ECONOMIC DEVELOPMENT	5
1.1 THE BENEFITS OF EVALUATION	5
1.2 INTRODUCING EVALUATION: HISTORY AND PURPOSE	10
1.3 METHODS AND THEIR ROOTS	17
1.4 EVALUATION TO STRENGTHEN SOCIO-ECONOMIC DEVELOPMENT	24
1.5 GOLDEN RULES	34
PART 2 DESIGNING AND IMPLEMENTING EVALUATION FOR SOCIO-ECONOMIC DEVELOPMENT	37
2.1 DESIGNING AND PLANNING YOUR EVALUATION	37
2.2 IMPLEMENTING AND MANAGING EVALUATIONS	55
2.4 GOLDEN RULES	80
PART 3 DEVELOPING CAPACITY FOR SOCIO-ECONOMIC EVALUATIONS	83
3.1 DEVELOPING INSTITUTIONAL CAPACITY	83
3.4 GOLDEN RULES	99
PART 4 CHOOSING METHODS, TECHNIQUES AND INDICATORS AND USING EVIDENCE IN EVALUATION	102
4.1 FACTORS INFLUENCING THE CHOICE OF METHOD, TECHNIQUES, DATA AND EVIDENCE	102
4.2 METHODS AND TECHNIQUES FOR EVALUATING DIFFERENT TYPES OF SOCIO-ECONOMIC INTERVENTIONS	108
4.3 METHODS AND TECHNIQUES FOR DIFFERENT EVALUATION PURPOSES	114
4.4 METHODS AND TECHNIQUES APPLICABLE AT DIFFERENT PROGRAMMES/POLICY STAGES	116
4.5 METHODS AND TECHNIQUES APPLICABLE TO DIFFERENT STAGES IN THE EVALUATION PROCESS	119
4.6 ACQUIRING AND USING DATA IN EVALUATION	122
4.7 CREATING INDICATORS AND INDICATOR SYSTEMS	126
4.8 USING INDICATORS TO IMPROVE MANAGEMENT	139
4.9 GOLDEN RULES	145

Annexes:

- Annex A The Main Stages of Evaluation
- Annex B Changes in Structural Fund regulations

INTRODUCTION

This GUIDE, is intended for those involved in the evaluation of socio-economic development in Europe. It is a successor to the MEANS collection (methods for evaluating structural policies). Whilst the GUIDE has a specific focus on evaluation within European Structural Funds, it is not confined to the evaluation of these interventions. Socio economic development is, after all, strongly featured in many national and regional programmes that are not funded by the EU. In most countries in Europe, as elsewhere in the world, improving regional and local economies that have fallen behind, re-integrating marginalized groups, and adjusting to the challenges of global competition and technical change are priorities. Given the scarcity of resources and the often innovative strategies that socio-economic development requires, the demand for evaluation has expanded alongside the development of policies and complex interventions themselves.

Evaluating socio-economic development; policies, programmes, themes and projects

The GUIDE is concerned with the evaluation of socio economic development in Europe which gives it a particular focus on European Structural Funds. The funds are organised in programmes and evaluation takes place at ex ante, mid term and ex post stages. The programmes are one means of achieving wider policy goals and programme evaluation contributes to policy evaluation. The programmes comprise many interventions and projects. Evaluation at the level of the measure/intervention/project forms a part of programme evaluation. Many different programmes and their elements contribute to thematic objectives and thematic evaluation builds upon project and programme 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 what are the strengths and limitations of evaluation 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. This sourcebook material is available via the Internet at: <http://www.evaled.info>.

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 Structural Fund requirements - has spurred many new publications in Europe. The European Commission has published detailed Methodological Guidance - on indicators, ex-ante evaluation, macro-economic evaluation etc - that is specific and closely aligned with the Structural Fund Regulations. There is also a Financial Regulation that requires ex ante and ex post evaluation of all EU funded programmes that has to be adhered to¹. 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 are bound to change. 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 intends to fill a gap in the market to broaden understandings of sound methods and good practice in an accessible form.

Updating MEANS

Of course the main source of such generic guidance up to now has been the MEANS collection - a valuable and comprehensive set of handbooks published by the European Commission in 1999. The MEANS collection has become a standard text for European evaluators and has enjoyed a justifiable reputation for its scope, ambition and coverage. Indeed many aspects of that collection have stood the test of time and have been incorporated into this new 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) and an emphasis on the institutionalisation of evaluation).
- European policy has moved on, especially following the Lisbon Agenda. 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 also poses challenges for evaluation. Structural Funds are being 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.

¹ Financial Regulation (Council Regulation 1605/2002), Article 27. Examples of evaluation rules and principles that must be noted when European funds are involved include Article 21 of the implementing rules (Commission Regulation 2342/2002) that, in particular, defines the scope of ex ante evaluation.

Together these changes are substantial and this GUIDE has been written 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 the GUIDE, Sourcebook material and Glossary are accessible and searchable via the Internet. In future it is intended that the material will be further developed so that users will be able to structure and configure their own 'personal' GUIDE that matches their needs, the evaluation role that they have and the particular programmes and policies that they are evaluating or managing.

Content and structure

The new 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 GUIDE itself is in four parts.

Part 1 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.

Part 2 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 evaluation work. 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 communications to ensure influence and managing quality assurance in evaluation.

Part 3 discusses how to develop evaluation capacity and strategies for capacity development are discussed. The argument is structures around an 'idealised' model that suggests four stages in capacity development. This part includes discussion of internal capacity within administrations, as well as external capacity within professional networks and partnerships.

Part 4 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: of 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 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.

Evaluating Socio Economic Development, The GUIDE: Introduction

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

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 – including the accession countries. It illustrates the advice provided in the GUIDE and is intended to stimulate the development of evaluation capacity.

Finally, there is a Glossary that contains definitions of the terms used in the GUIDE and Sourcebooks.

PART 1 THE CONTRIBUTION OF EVALUATION TO SOCIO-ECONOMIC DEVELOPMENT

This first part of the GUIDE begins with a reminder of why evaluation is so important in socio-economic development programmes. It introduces some basic evaluation ideas, their origins and purposes. It then highlights some of the characteristics of socio economic policies and interventions and the implications of these for how they can be evaluated.

1.1 THE BENEFITS OF EVALUATION

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 regeneration of inner cities, 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 Structural Funds 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 programme stakeholders – whether they are managers, policy makers or beneficiaries.

Providing answers to worthwhile questions

The contribution of programme 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.

Evaluation can reduce uncertainty and improve planning and implementation

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

Evaluation capacity and integration into the policy cycle

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?

Balancing the pragmatic with the 'state of the art'

Improving programmes over time

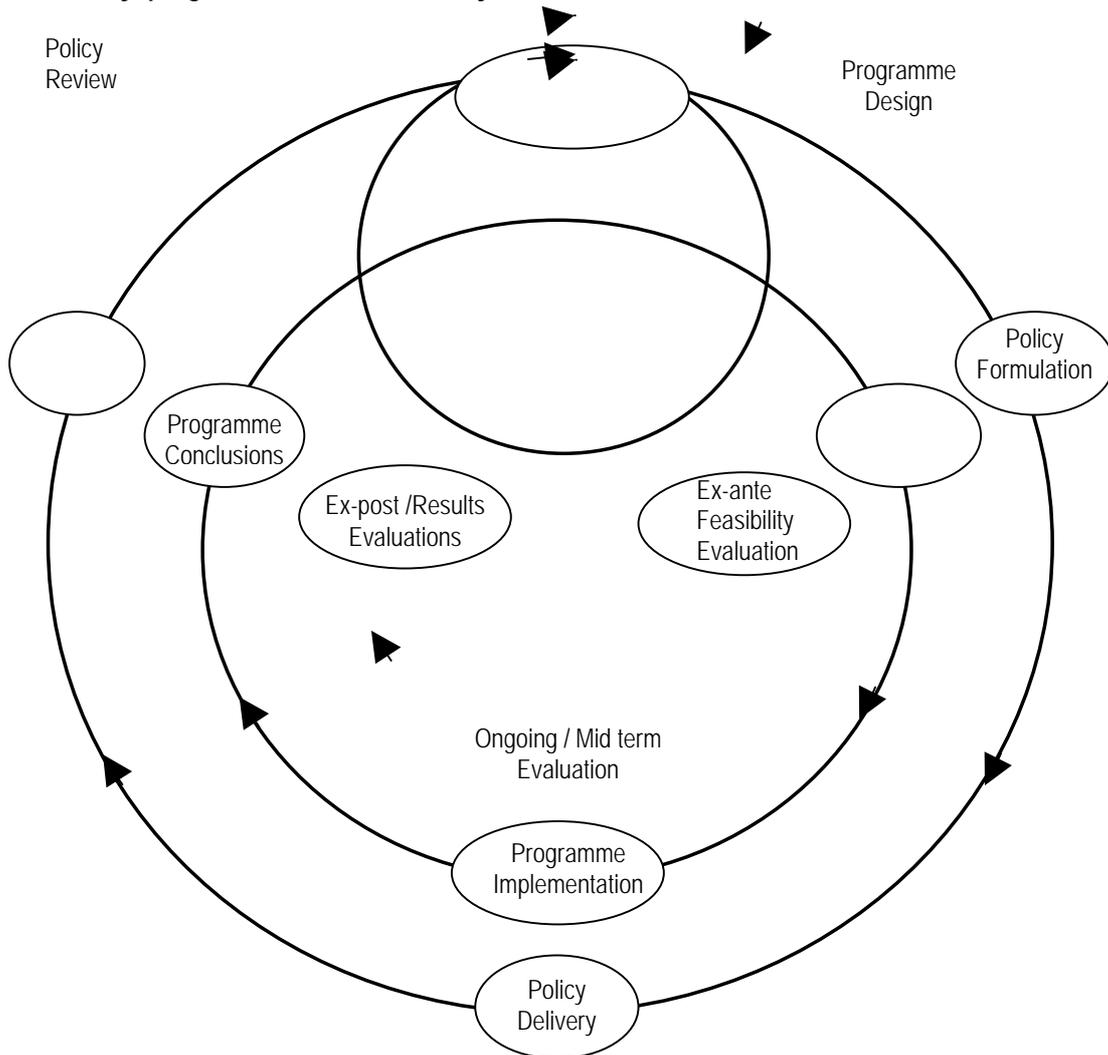
One important organising principle that runs through the GUIDE is the time-line of policy. It is common to speak of the 'policy cycle' that begins when policy (and associated programmes) are formulated and continues through planning and resource allocation, programme design, implementation and the delivery 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 Structural Funds. These terms are elaborated in Annex A. A similar logic is present in the distinction often made between 'outputs', 'outcomes', 'results' and 'impacts'.

Different stages of a policy and programming cycle

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 Box1.1 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.

Box 1.1 Policy, programme and evaluation cycles



Box 1.1 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 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 – not unusual in national and European programmes of socio-economic development. This can, for example, lead to changes in objectives or priorities after systems

Keeping these cycles aligned

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.

In general we are interested in this GUIDE in the evaluation of programmes rather than the evaluation of policies. However these are sometimes not easy to separate. This is especially so in an age of what is sometimes called evidence based policy. There is a strong imperative to use resources wisely and to learn from previous experience: to base policies on evidence. This is indeed a key feature of the way ex ante evaluation is understood in European Structural Funds. Evaluation can be a powerful tool for extracting lessons learned from programmes so as to improve policies in the future.

*Learning for
policy purposes*

There are a number of particular contributions that evaluation can make throughout the programme cycle:

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.

*Ensuring the
relevance of
programmes*

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.

*Assessing the
strength and
eligibility of
interventions
and instruments*

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

Early 'results'

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.

and processes

Identifying outputs, outcomes 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', eg 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 'outcomes' are expected to have a bigger impact that relates to policy and programme goals – at least 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?

Substantial results and outcomes

For many policy makers, identifying, describing and quantifying such outputs, outcomes 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.

Identifying unintended consequences and 'perverse' effects

Even when programmes and instruments fulfil their stated objectives these 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 unanticipated consequences and 'perverse' effects is essential. This is also a way in which evaluation can contribute to learning lessons – in this case mainly learning interventions, how to design programmes better and how to avoid wasteful interventions and 'perverse' effects.

Capturing unintended consequences is important for learning

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.

Linking interventions programmes and policies is often a problem

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.

...especially as there is a tendencies to move evaluation upstream to the policy level

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 main 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.

Self evaluation can help at the intervention /project level

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.

1.2 INTRODUCING EVALUATION: HISTORY AND PURPOSE

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 (eg the effectiveness of new curricula in schools); linking evaluation with resource allocation (eg through a Planning, Programming and Budgeting system) and the evaluation of anti-poverty programmes (eg 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 – the basis for many educational testing experiments; cost benefit and

Different strands of evaluation tradition

economic appraisal methods; and participatory and qualitative 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, eg elected officials and politicians;
- professional and specialist interests, eg teachers in education or scientists in research;
- managers and administrators, eg civil servants and managers of local public agencies;
- citizens and those affected by public action, eg 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.

Different stakeholders want evaluation to help them in different ways

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 has 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.

There are different national traditions

European Structural Funds have been a major driver for spreading the practice of evaluation throughout the EU. At every stage of the programming cycle (ex-ante, mid-term, 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.

Structural Fund approach to evaluation has evolved over time.

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, 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 (see Box 1.2).

There have also been transitions in the development of evaluation regulations

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 principal. The range of potential stakeholders in evaluation has therefore expanded to include, for example, local authorities, social partners and civil society groups.

Shifting responsibilities for evaluation

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, requires that these evaluations are used both at the ex-ante stage, and again at the mid-term. The revision of the mid term evaluations are termed final evaluations. This combination of devolved responsibilities with external scrutiny by higher tiers of government is also typical of national evaluations of socio-economic development.

Box 1.2 The Evolution of Structural Fund Regulations

1994 1999	2000 2006
EX-ANTE EVALUATION	
<p>An <i>ex-ante</i> 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 <i>the ex-ante</i> evaluation</p>	<p>The Member State has primary responsibility for the <i>ex-ante</i> evaluation; The aim of the <i>ex-ante</i> 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 <i>ex-ante</i> evaluation is incorporated in the plans</p>
MID-TERM EVALUATION	
<p>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).</p>	<p>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 mid-term 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 <i>ex-post</i> evaluations, and must provide the information required for the <i>ex-post</i> 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).</p>
EX-POST EVALUATION	
<p>An <i>ex-post</i> 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.</p>	<p>The Commission has primary responsibility for the <i>ex-post</i> evaluation in partnership with the Member State; The objective of the <i>ex-post</i> evaluation is defined; The <i>ex-post</i> evaluation is carried out by an independent evaluator within three years of the end of the programming period.</p>
PERFORMANCE RESERVE	

Before 31 March 2004, the 4% of the indicative allocation for each Member State held back at the beginning of the period is allocated as a 'performance reserve' 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, effectiveness, management and financial implementation which are defined by the Member State in consultation with the Commission.

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 (see *Sound and Efficient Management - SEM 2000; and Spending more Wisely – Implementation of the Commission's Evaluation Policy*). This has been taken further most recently by linking financial and budgetary decisions to performance that therefore needs to be measured and described. Performance management in the European Commission (see *Implementing Activity-Based Management in the Commission*) as well as in Member States is now reflected in a further element within the evaluation obligations of the current regulations, those that concern the Performance Reserve. Indicators of effectiveness management and financial implementation have to be gathered in order to justify the release of funds held back at the beginning of the programming period.

The link with performance management

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.

Many roots and different origins

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

Four sources of ideas and theory

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 in contemporary political perspectives highlights 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 (this is taken further in Part 2 of the GUIDE when we discuss the 'profile' of the evaluation team and choosing the right evaluators, page 55).

Bridging
diversity within
an evaluation
team

At a systemic level (eg 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 (this is also taken further in Part 2 of the GUIDE).

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.

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 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 strengthening* - improving and developing capacity among programme participants and their networks and institutions.

Five different purposes

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 an 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

Different purposes interest different stakeholders

..... and pose different evaluation questions

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?

Learning as an overarching 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.

We need to learn how to improve socio-economic development

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 below. 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.

1.3 METHODS AND THEIR ROOTS

Part 4 of the GUIDE 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 Part 4 of the GUIDE and are introduced here.

The methodological axis

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.

*Four
methodological
'positions'*

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 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.

*Different types
of evaluation
allow us to do
different things*

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

*The challenge
of evaluating
complex
programmes*

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.

Each of these policy areas or domains brings with them their 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.

Each policy area has its own evaluation traditions and assumptions

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'.

Socio-economic development varies across settings – common laws are difficult to find across contexts

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.

It is stakeholder understandings that are important

Opening the black box to establish causal mechanisms

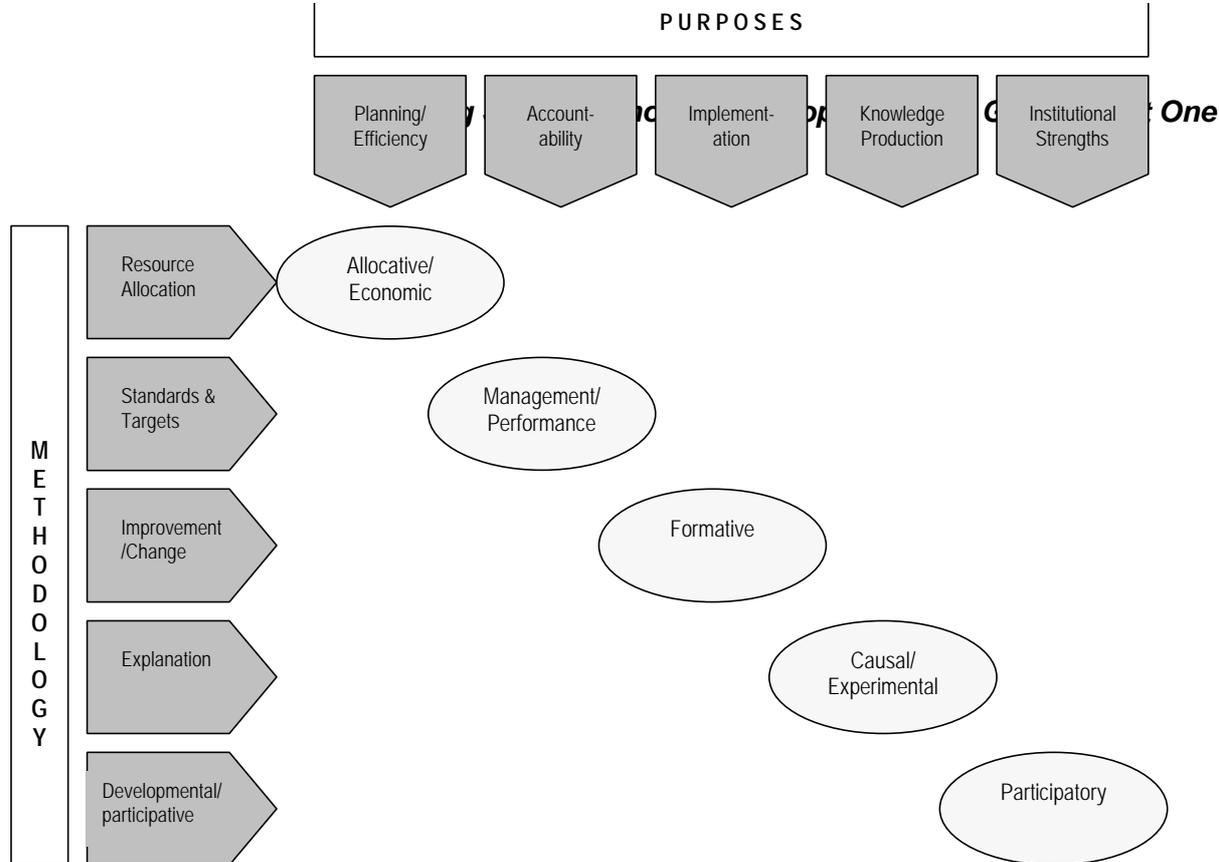
In practice evaluators are unlikely to see themselves as operating exclusively within any one of these philosophical traditions but will tend towards one or another depending on the circumstances of the

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 in Box 1.3. 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 reflect elements of several types of evaluation.

Box 1.3 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.

Five main types of evaluation

These 'types' cannot be rigidly located – even though it helps to place them in an approximate way. For example formative evaluation can also help strengthen institutions and help management achieve targets and allocative/economic evaluations can contribute to accountability 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 tools and techniques in Part 4 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. Many of the evaluations required under European Structural Funds fall into the former category – although there is also an expectation that ex ante and mid-term evaluations will contribute to programme improvements.

However, many evaluations in socio-economic development contexts are entirely about advice, development and improvement. 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.

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 so 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

*Good theory
makes
evaluation
practical*

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, ie 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, ie how does development occur spatially and sectorally (see Section 1.4 below). There are similar bodies of theory in most policy priority areas eg in education; health; employment; environmental protection etc. Sourcebook 1 elaborates on the theories linked to each of the themes and policy areas.

Evaluators should have knowledge of policy area theory as well as evaluation theory

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.

1.4 EVALUATION TO STRENGTHEN SOCIO-ECONOMIC DEVELOPMENT

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, crime prevention 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.

Despite many different interventions some generalisations are possible

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

certainly in Structural Funds interventions are 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.

The characteristics of socio-economic development ...

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 what should be the evaluation focus.

.....has consequences for how we evaluate

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

- 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 generally poor. These circumstances can be compared with programmes where

A need for structural change in the face of persistent development problems

interventions are less ambitious, for example, where they involve introducing a new qualification system or upgrading an existing road system.

- 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. *Multi-interventions for multi-dimensional developments*
- 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. *Matched to the characteristics of settings*
- 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 the European Structural Funds, shaped by a general commitment to socio-economic cohesion through the reduction of disparities of GDP per head and other conditions across the EU. This would also be true of many national and regional socio-economic programmes. *Within a broader policy framework*
- 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. Indeed in many socio-economic programmes that adopt a local development approach, such partners take the lead in setting priorities, designing programmes and implementing and eventually managing programme outputs. *Importance of bottom-up partnerships*
- 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. *Longer term development and sustainability*

We can directly link these characteristics of socio-economic development with evaluation. Some of the main links are summarised in the Box 1.4 below.

Box 1.4 Implications for evaluation of the characteristics of socio-economic development programmes

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 (eg 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

Box 1.4 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

Evaluating in a partnership context

Structural Fund programmes 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 Part 2 of the GUIDE. There are also methods that can be deployed to integrate the multiple criteria of different stakeholders (see Part 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.

The EU policy context

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

Structural Funds programmes have their own orientations and particular horizontal priorities. The horizontal priorities for the 2000-2006 period are: employment and human resources; sustainable development; environment; equal opportunities for men and women; Information Society; and Local Development. Structural Funds are also embedded within a wider European policy agenda, which increasingly shapes the content of programmes.

Horizontal priorities for European development programmes

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' (CEC 2001). 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

Reducing regional disparities

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** 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. The links between these various policies and Structural Funds are most evident in relation to the new Member States, for example environment measures related to the sustainable development strategy endorsed by the Göteborg European Council, and human resources priorities linked specifically to the Lisbon agenda are highlighted in guidance offered by the European Commission (Communication from the Commission: Further Indicative Guidelines for the Candidate Countries COM2003 110 final). Even though there may be differences in emphasis between those 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.

Combining economic and social dimensions

Evaluations of socio-economic development therefore 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:

The policy challenges for evaluation

- Many of the policies and programmes (eg 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 combine 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 the European Structural Funds and other socio economic development interventions, horizontal priorities are further reinforced by cross-cutting policies, 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 how the component parts interact with each other is part of the complexity that characterises socio-economic development programmes.
- Partly 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 measure 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 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 particular 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 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.

The contribution of programmes to policy evaluation

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 *From demand*

managing demand through regional subsidies and other subventions (eg payments to the unemployed), current thinking is more directed at supply or capacity. This can take various forms – mobilizing underused resources, increasing the capacity and value of existing resources and ‘transferring’ new resources into a region or sector. *to supply and capacity*

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. *‘Know-how’ and its application as a product and contributing to production*
- **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 its human capital into greater economic prosperity. *Productivity of the workforce*
- **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. *Supporting effective human relations*
- **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. *Disparities in access to goods, services and labour market*
- **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. *Imitation reduces disparities*

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, culture – often the basis for tourism). *Exploiting that which is distinctive*

- *Growth 'poles'*. That growth may require at first 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. *Concentration before disparities reduce*

These different theories that underpin policy priorities such as cohesion, the European Employment Strategy 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. *Why additionality matters*
- 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. *Are programme resources targeted?*

In many cases the ways in which programme evaluation contributes to answering big policy questions is through a qualitative understanding of what is really going on '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 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. *Programme evaluations help*
- Assess the quality of these investments in terms for example of their relevance, uptake and perceived usefulness by local and regional actors. *understand what goes on 'on the ground'*
- 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. *Help programme managers*
- Reassure policy makers that monies are being spent for the purposes *Account for*

for which they were made available – well ahead of programme completion. *public money*

- Provide contextual information that will allow policy evaluators who are assessing the effectiveness of policy instruments more generally to interpret their findings. *Provide context for policy evaluation*

Community Value Added

Although the GUIDE is not exclusively concerned with the evaluation of European Structural Fund interventions much of the advice is pertinent to these programmes. Often the evaluations of these programmes 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 additional to what would otherwise 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 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 mainstreaming that are a consequence of EU financing of the intervention should be assessed. (In this respect as discussed in Part 3 of the GUIDE 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.

To provide a balanced assessment these and other aspects of Community value added should be set against any attendant ‘transaction costs’.

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 have 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

An ideal world...

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 often not like that! Policy priorities change or evolve whilst 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.

...that rarely exists

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.

Producing what is possible...

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 Part 3 of this GUIDE and in the associated Sourcebook 3. Of course, there will be some circumstances where the conditions are so poor that it would be unwise to conduct an evaluation. Such a conclusion might be reached on the basis of an 'evaluability assessment' (see Sourcebook 2). Arguably in such circumstances the wisdom of continuing with the programme can be questioned.

Whilst developing capacity for the future

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.

1.5 GOLDEN RULES

This part of the GUIDE has introduced some of the main issues 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 or 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 eg 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 and/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 evaluations 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 Part 4). 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 being constantly 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. The broader criteria that need to be designed in to evaluations usually derive from the wider policy framework.
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, 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 so 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 is 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.

PART 2 DESIGNING AND IMPLEMENTING EVALUATION FOR SOCIO-ECONOMIC DEVELOPMENT

This second part of the GUIDE provides advice on the design and implementation of evaluation for socio-economic development. It considers in turn: the designing and planning of evaluation; .and the implementation and management of evaluation assignments.

2.1 DESIGNING AND PLANNING YOUR EVALUATION

Evaluation and programming

In part 1 of this GUIDE 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 if not downright incompatible.

Planning documents may be vague and inconsistent

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/or 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 for 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 – see *In the early stages,*

Part 4 and Sourcebook 2) can be used, from a very early stage, in order to clarify the starting priorities and the logic that underpins programme interventions.

evaluation helps to clarify the baseline and priorities for action.

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 set up that can ensure that the process is kept on track throughout the programme’s life.

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/or of 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?

The value of the evaluation, more often than not, lies in posing the right questions as well as in providing the correct answers.

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.

Evaluation is easier when the programmes logic is explicit

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.

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.

Whatever the origin of objectives they provide one important set of criteria against which the achievements of interventions can be judged.

Planning evaluation work

This section more precisely discusses the various activities and issues that are needed when planning evaluation work. 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 the limits of the programme 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

Expectations for evaluation will vary at the different stages – ex ante, mid term and ex post.

programming cycle under consideration or to a certain extent to include funding of preceding cycles?

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 and its most essential interdependencies with other public policies and interventions.

Scope refers to what is being evaluated.

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 – a likely output from developing 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. This is something that is likely to be a key output of the inception phase.

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 to the programme 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.

There are many stakeholders all of whom have to be

Other factors which have reinforced the trend towards involvement of a 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.

*involved in
some way*

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 program formulation process (the “bottom up” approach to planning) an essential step of the whole exercise.

*...especially in
bottom up
local
development
settings*

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 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.

*Involvement
also
encourages
ownership of
evaluation
results and
findings*

The first question that must be asked, after that 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/or can be interested in the process or in the results of the evaluation itself? This phase in the evaluation parlance is called ‘the identification of the stakeholders’.

*Identify the
‘affected’ or
potentially
affected actors*

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 even in the case in which 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 are to provide evidence (data, information, judgements, etc.) as part of the evaluation process. Many methods and techniques described in Sourcebook 2 can be used to this purpose: individual interviews, surveys, focus groups, etc.
- At the other end of the continuum the stakeholders can be involved in the steering of 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 will discuss the management of the evaluation process.

Stakeholders should input to the evaluation design.

In practice the involvement of the stakeholders in most programmes 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 practised through the dissemination of reports, ad hoc meetings and similar instruments.

Programme theories and logic models

Stakeholder consultation phase also provides an opportunity to reconstruct the logic of the programme prior to its launching. As we have seen in Part 1 of the Guide, there are different and often competing theories underpinning interventions in this field.

Every policy and programme should state its underlying assumptions, supported by evidence

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 should be consistent with each other and should 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. Part 4 of this GUIDE (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 the 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.

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?)

Evaluation questions can be descriptive causal, normative, predictive or critical

Ideally, evaluation questions should have the following qualities:

- The question must correspond to a real need for information, understanding and/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.

Evaluation questions should relate to decision-making, concern impacts or needs, and include judgement criteria

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 efforts but can be addressed through existing monitoring systems, consulting managers, or referring to audit or other control systems.

Evaluation criteria

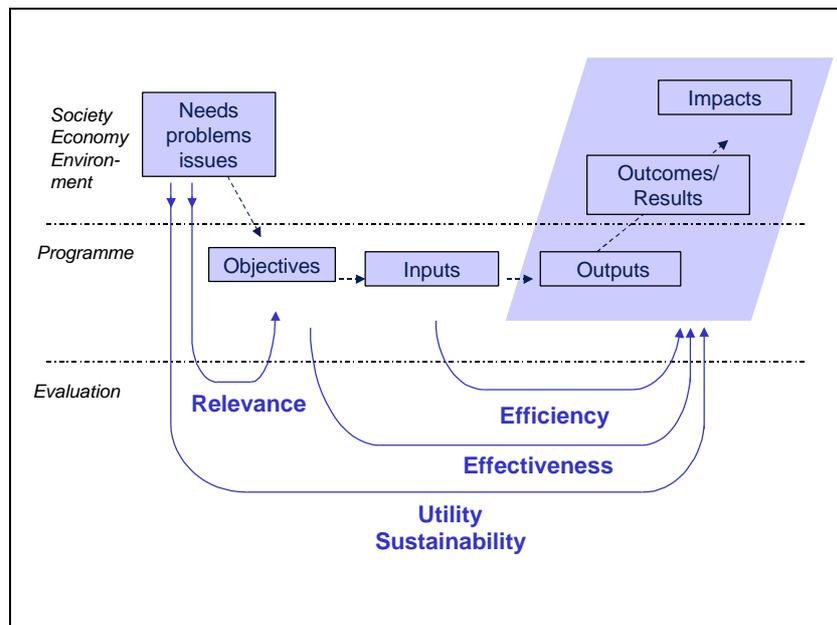
Evaluation questions that include judgement criteria fall primarily into one of the following four categories:

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

The main judgement criteria are concerned with relevance, effectiveness, efficiency and utility

These four main categories are represented in Box 2.2.

Box 2.1 Main evaluation criteria



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 the choosing the best strategy or on 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.

Relevance denotes whether the intervention was appropriate to the problems it sought to address

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.

Effectiveness refers to the achievement of

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 in much the same way as efficiency).

objectives....

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.

...and efficiency takes account of the associated costs

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.

Utility judges the impacts against the wider social and economic needs

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, as in this GUIDE and the accompanying Sourcebook 1, 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 in Box 2.2.

Box 2.2 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.

A wide range of other criteria could also be used to make evaluative judgements

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 resources. Will the evaluation team, with the available time and resources

The 'evaluability' of the questions is the extent to

and using appropriate evaluation tools, be able to provide credible answers to the questions asked? This requires an 'evaluability' study should be carried out.

which they can actually be applied in practice

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

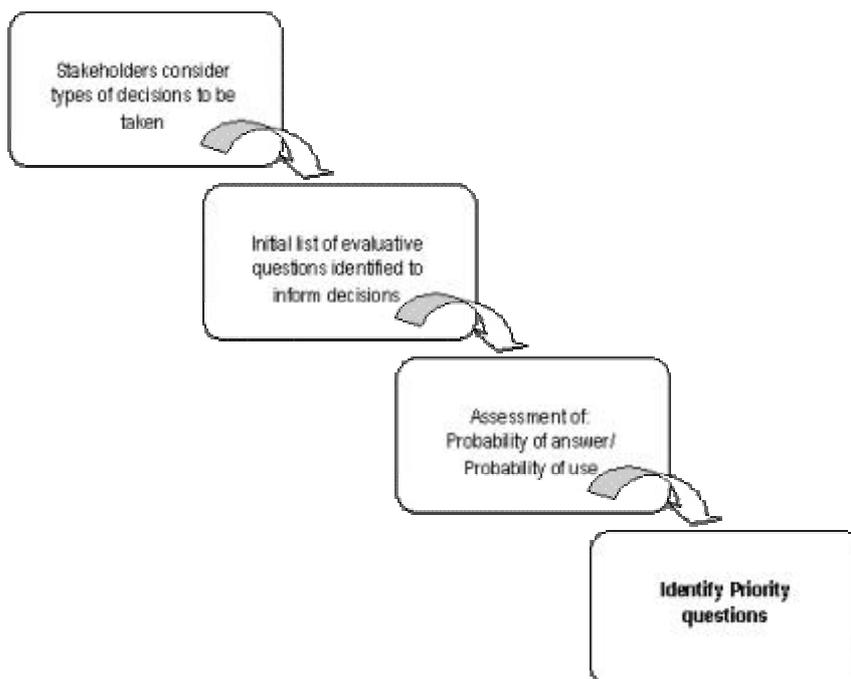
- whether the concepts are stable,
- whether explanatory hypotheses can be formulated,
- whether available data can be used to answer the question, without any further investigation,
- whether access to the field will pose major problems.

*Questions should be graded according to usefulness of information obtained
Assessing evaluation questions helps save time and effort*

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.

As Box 2.3 indicates, important considerations at the evaluability stage are the probabilities that evaluation results will be obtained and used.

Box 2.3 Selecting priority evaluation questions



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.

Choosing evaluation methods and 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.

The influences on the choice of methods are varied

Part 4 of this Guide, 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 facts that the evaluations in the Structural Funds are often compulsory and that both the European Commission and the National authorities in charge of the funds issue guidelines about when and how to perform the evaluative exercises is a mixed blessing. On the one hand it tends to routinise the decision to evaluate. Evaluation becomes only 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 Part 3.

Evaluations of Structural Fund programmes are compulsory.

Guidelines are especially useful to set the parameters for evaluation in relatively de-centralized 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.

Key Resource decisions

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

*Top
management
and key
external
partners*

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:

*should be
involved in
planning and
evaluation*

- 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 evaluative study, 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.

Ideally the people in charge of the evaluation should have some practical experience of evaluation.

Those in charge must be senior enough to have direct access to the policy makers at all levels in order to share with them the knowledge that the study is supposed to 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 to the evaluation team the right questions.

Ideally, therefore, the people in charge of the evaluation should have some experience of the practical work of the 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.

Spend more for innovative programmes and interventions

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 commitments 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 evaluation work that has 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 evaluation 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.

In house and external teams each have their advantages

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.

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 clearly 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 Terms of Reference set out the basis of the assignment

The ToR should be brief (typically 5-10 pages) supplemented if necessary by administrative annexes. A model content for a ToR is listed in Box 2.5 and is then elaborated.

Box 2.5 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. Whether an evaluation is obligatory and legally required, or is undertaken because programme managers see it as important should be made clear.

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've 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 is intending to ask. To focus the evaluation over a narrow list of questions that are relevant for the commissioner ensures a 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 considerable 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, eg 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 suggestions 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 proposing 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've 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;
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 Structural Funds, more and more 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.

Choosing evaluators on the basis their skills, expertise and prior knowledge

The overall choice should depend less on the institutional origins of the evaluation team and more on the required competencies, ie 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 in order to provide guidance to the potential applicants, the ToR should specify how the proposal has to be structured, possibly indicating the maximum number of pages to 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 ToRs 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 selection criteria are applied systematically once proposals are received.

2.2 IMPLEMENTING AND MANAGING EVALUATIONS

Choosing the right evaluators

The process of selecting evaluators needs to be transparent. Careful drafting of the ToR is the best possible way to attain this goal, together with the use of a formal selection committee. This should not only include representatives of the people in charge of the evaluation, but also an independent expert and, when possible, representatives of the potential and actual users of the evaluation.

Carefully drafted Terms of Reference and a balanced selection

Members of the selection committee should reach their own judgements on tenders against the criteria given in the ToR. The judgments 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.

*committee help
to make the
selection
process
transparent*

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 Box 2.6.

Box 2.6 Assessing the quality of the method of approach in a proposal

For each evaluative question

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

*The method
should be
suited to the
evaluation
questions
involved*

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 and/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 fulfil the specifications in the ToRs. It could be argued that if the evaluation is standard and/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.

*Capacity to
deliver the
evaluation is
paramount*

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/or 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 – ie 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 pertinent.

Basing judgements on previous track record can discriminate against new entrants to the evaluation market

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 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.

The costs of the evaluation should be only one element in the decision

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.

Managing the evaluation process

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

Minimising contact between the commissioning authority and the evaluation team will not ensure independence of the evaluation...

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 work 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 of 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' on evaluation underway 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 to 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 activity is normally both summative and formative and the balance between the two may well shift during the implementation of evaluation work. If those commissioning evaluation work 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 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 partners 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 later in this part of the GUIDE, 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 it 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 work 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 work 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. Here the intervention is relatively small and complex but involves different parties perhaps working together for the first time and the evaluator may be as much an 'animateur' and catalyst for consensus as impartial observer. Very often evaluation work 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 where 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 mid-term, 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.

...in fact, close links between the parties are to be preferred

Box 2.7 Ex-ante evaluation in Czech Republic

In the context of the ex-ante evaluation of the National Development Plan in the Czech Republic it was noted that the recommendations of the evaluation team proved to be more readily accepted if they were suggested in early stages of drafting programming documents. This implies that evaluation was perceived as more useful when real partnership were established and in any case the work of evaluators had a lot of features that are usually associated with technical assistance.

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 therefore 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;

An interactive approach can be ensured through regular Steering Committee meetings

- 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, and subsequently continuously updated. 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.

The Inception Report provides the main point of reference for quality assurance

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 Structural Funds, this 'interim' stage is often included in the mid-term evaluation. This emphasises one of the advantages of ensuring that mid-term and ex post evaluations are linked. 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 mid-term 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 body such as 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 Structural Funds shows the advantage of the involvement of the most important stakeholders, and in particular the relevant institutional and other key partners – those actors whose co-operation is needed in order to bring about the main results of the programme. The advantages of an 'inclusive' Steering Committee are shown in Box 2.8.

Box 2.8 The advantages of an inclusive Steering Committee

Establishing an evaluation Steering Committee consisting of the different stakeholders in the programme makes it possible to guarantee:

- 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 likelihood 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, that is 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 very important as programmes grow increasingly complex taking into account concerns that have different territorial dimensions;
- *The operational management* of the programme, that is those whose activities are evaluated by the study, 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 of 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*: ie the people representing the main interests affected by the programme. These can include not only the trade associations, trade unions and the economic interest associations, but also the institutional and/or 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 be very important in providing useful inputs to the evaluation team and in order to open up debate towards the more general lessons that can and should be drawn from the exercise.

The Steering Committee should cover a range of interests

The principle 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 evaluating questions and key issues that they feel should be informed. 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.

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 understandings rather than put all communication efforts into one big dissemination exercise after a final report is delivered. Communication should therefore include:

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

Improving awareness of the evaluation work that is 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.

Raising the awareness of evaluations

Providing feedback on interim findings

The communications of interim findings pose 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. It may not be appropriate to have no formal Interim Report. This can avoid the criticism that partial and unsubstantiated evaluation findings can attract. At best 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. Such findings which can express findings as 'issues' and 'pointers' can fulfil an important learning purpose in evaluation terms.

Early findings contribute to learning

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

Production of report that is acceptable to the evaluation team, commissioning authorities and

be insensitive to the political import of what they present – especially at draft stage. 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 can be a major challenge and require a good deal of time. The following suggestions may facilitate the process: *stakeholders*

- 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 managing authority should take responsibility for the circulation of the report and compiling feedback.

Communicating evaluation findings

Evaluation work 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 practice lessons:

Evaluation is wasted without communication of findings

- 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 evaluators' 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. The careful identification of potential users, from policy makers through beneficiaries to the general public, need to be identified and the most appropriate channels of communication selected.

Accessible distribution methods should be tried

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.

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 and/or from the flaws of the planning and design phase. Unfortunately those involved in these three sets of activities are different, and very 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 can be considered a *characteristic of the products* of the evaluation process. Thus one could analyse the ToR

Assessing the quality of evaluation is complex, but generally relates to the planning and design, implementation and the quality of data

Quality relates to the processes and the products of evaluation

according to the criteria that we have already spelled out. More likely, 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.

Good practice standards have been developed in recent years

Since then, however, some things have improved. In particular, as Box 2.9 shows, it is now becoming common to attempt 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).

Box 2.9 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 – rather than technical practice of evaluation. But again this mainly concerns the ethics of evaluators rather 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.

Many have been developed by evaluation societies...

...some are more concerned with ethics rather than technical practice

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.

Box 2.9 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 für Evaluation (DeGEval): Standards für Evaluation (2001)
<http://www.degeval.de/standards/index.htm>

Société canadienne d'évaluation (SCÉ) Lignes directrices en matière d'éthique / Guideline for Ethical Conduct
<http://www.evaluationcanada.ca/>

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/EvaluationDocuments/aeaprin6.html>

Australasian Evaluation Society (AES), Guidelines for the Ethical Conduct of Evaluations
http://www.aes.asn.au/ethics_guidelines-1.pdf

UK Guidelines for good practice
http://www.evaluation.org.uk/ukes_new/Pub_library/GuidanceGoodPractice.doc

PUMA Best Practice Guidelines
<http://appli1.oecd.org/puma/bpi/bpiste.nsf/pages/Evaluation>

Italy Treasury Guidelines
http://www.dps.tesoro.it/documentazione/docs/all/Criteri_qualita_sistema_nazionale_valutazione_maggio2002.pdf

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 (Box 2.10).

Quality assurance and Quality control

Box 2.10 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).

Box 2.11 identifies both quality control and quality assurance criteria. Both are needed as a means of judging evaluation reports and outputs. Normally the person responsible for the 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. The quality control and quality assurance criteria are elaborated in Box 2.11.

Box 2.11 Judging evaluation reports and outputs

**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 in which programme objectives have evolved been interpreted and 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?

Answering the questions posed in the Terms of Reference

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

The scope of an evaluation must cover questions that are relevant from the point of view of the programmes.

Answering questions that are relevant to

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 have been well described and whether the problems and successes in the implementation of the programme have been properly clarified.

the essential characteristics of the programme

Secondly, because the results and impacts of the programme have to be analysed in order to judge the extent to which programme 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 not 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 target that can be defined in terms of eligible geographical areas or non-localised target groups (eg the long-term unemployed). It is therefore necessary to check whether:

- the limits of the scope, in terms of areas or social 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 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?

Appropriate methodological choices

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 whether:

- 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.

Sources and methods provide reliable data

Sound analysis

Quantitative analysis consists of the analysis of data in the form of tables or any other form of statistical analysis. Qualitative analysis consists of the systematic comparison and interpretation of information sources in the form of cross-referencing. 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 according to instructions in the relevant technical manuals.

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.

The quantitative and qualitative analysis underlying the analysis and conclusions is sound

For this reason a before-after comparison or, when similar groups do not exist, a comparison between beneficiaries and a control group, is recommended. In the former case this before-after comparison must be carried out appropriately, otherwise it must be established whether that would have been possible. In the latter case, the comparative analysis must be able to rely on the collection of data within similar or control

groups.

Credible results

The credibility of results is defined here as the fact 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.

Results that are justified by the evidence

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:

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

Impartial conclusions

Conclusions include suggestions and sometimes even 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:

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

Conclusions are more than results

Key questions such as relevance, effectiveness and efficiency of the programme 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 programme budget, both globally and in terms of internal allocation to the different axes and measures.

Essential questions such as the value added of the programme and progress made in terms of transversal goals like cohesion, subsidiarity, good governance, sustainable development and equal opportunities need to be studied. In the case of ex ante exercises, 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.

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 among others, and continual communication of results is desirable. The legibility 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 whether:

- the report was written clearly and it 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;
- 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.

Clarity of reporting

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 the future in order to improve the quality of future evaluations.

Coherent and evaluable objectives

The coherence of the programme objectives: the extent to which they are specific, linked to interventions, not contradictory etc. has been discussed earlier. It was also noted that the use of logic models, programme theory and theory of change approaches are useful ways of clarifying 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 in terms of objectives and interventions. 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?

However the purpose of these opportunities for briefing and exchange is the dialogue and feedback that they enable. Was there a 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. Part 4 of this GUIDE 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 this is so. 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 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 – say 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 itself 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 workplan 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 workplan 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 by organisations and agencies. Other civil society, commercial and informal groups also have an interest in evaluation findings - whether as taxpayers, local electorates or potential beneficiaries of a programme and its interventions. 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 not all the responsibility rests with evaluators. Programme managers and those who commission evaluations should also 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.

*Qualification
and
contextual-
isation of
quality criteria*

Box 2.11 indicates a grid for the quality control of the evaluation report. Box 2.12 provides a quality assurance grid.

In both cases 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.

Box 2.11 Grid for an assessment of the quality of an evaluation report

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:	<input type="checkbox"/>					
2. Relevant scope: The rationale of the programme, its outputs, results, impacts, interactions with other policies and unexpected effects have been carefully studied:	<input type="checkbox"/>					
3. Open process: The interested parties – both 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:	<input type="checkbox"/>					
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:	<input type="checkbox"/>					
5. Reliable data: The primary and secondary data collected or selected are suitable and reliable in terms of the expected use:	<input type="checkbox"/>					
6. Sound analysis: Quantitative and qualitative data are analysed in accordance with established conventions, and in ways appropriate to answer the evaluation questions correctly:	<input type="checkbox"/>					
7. Credible results: The results are logical and justified by the analysis of data and by suitable interpretations and hypotheses:	<input type="checkbox"/>					
8. Impartial conclusions: The conclusions are justified and unbiased:	<input type="checkbox"/>					
9. Clear report: The report describes the context and objectives, as well as the organisation and results of the programme in such a way that the information provided is easily understood:	<input type="checkbox"/>					
10. Useful recommendations: The report provides recommendations that are useful to stakeholders and are detailed enough to be implemented:	<input type="checkbox"/>					
In view of the contextual constraints bearing on the evaluation, the evaluation report is considered to be:	<input type="checkbox"/>					

Box 2.12 Grid for an assessment of the quality of the evaluation process

Please assess the quality of the evaluation process in terms of how positively or negatively it met each criterion specified below:

Very positive

Very negative

1. Coherent Objectives and Programme: The programme objectives were coherent and the programme was able to be evaluated:	<input type="checkbox"/>				
2. Adequate Terms of Reference: The ToR were well thought up and proved useful and did not need to be revised:	<input type="checkbox"/>				
3. Tender Selection: This was well conducted and the chosen tenderer was able to undertake the evaluation to a good standard:	<input type="checkbox"/>				
4. Effective Dialogue and Feedback: An inclusive and process was created that provided feedback and dialogue opportunities with Commissioners and managers that improved the quality of the evaluation:	<input type="checkbox"/>				
5. Adequate Information: Required monitoring and data systems existed and were made available/accessed by administrations and partners:	<input type="checkbox"/>				
6. Good Management: The evaluation team was well managed and supported by programme managers:	<input type="checkbox"/>				
7. Effective Dissemination to Commissioners: The reports/outputs of the evaluation were disseminated to Commissioners including Steering Committee members and programme managements who responded appropriately with timely feedback/comments:	<input type="checkbox"/>				
8. Effective Dissemination to Stakeholders: The reports/outputs of the evaluation were suitably disseminated to all stakeholders and where necessary targeted in ways that supported learning lessons:	<input type="checkbox"/>				

2.3 THE USE OF EVALUATION FINDINGS AND KNOWLEDGE MANAGEMENT

Undertaking evaluation work 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 work is used.

- Individual evaluations may be used directly or in an 'instrumental' manner whereby 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 individual evaluations combined with other evidence and opinion are used cumulatively to inform debates and influence decision-making. Evaluation work 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 work. 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 work. 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 work. 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 Part 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 evaluation work 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.

2.4 GOLDEN RULES

1. Evaluation competence should be brought in early by programme planners. In particular, this can help clarifying objectives and the intervention logics of programmes. This activity, although employing evaluation competence, is quite separate from mainstream evaluation activities. It needs to occur at the programme 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 main evaluation, 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 programme 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 these outputs. On the other hand, it may be necessary to be selective in 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 – 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, ie 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 the programme 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. Whilst this is a technical question – and this part of the GUIDE 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 that managers have. 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 an 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 even 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 clarify their own thinking and understanding of the programme and spark off innovative ideas for programme improvement. 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 part of the GUIDE 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. Individual 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.

PART 3 DEVELOPING CAPACITY FOR SOCIO-ECONOMIC EVALUATIONS

This third part of the GUIDE discusses some of the pre-conditions for the evaluation of socio-economic development - how to develop the capacity to undertake evaluations. It begins by outlining the nature of the capacity problem in terms of the needs of public management and in terms of the supply and demand side of the capacity equation. It then presents an idealised model of capacity building that indicates some of the stages in capacity building and how it can be achieved. The contents of this part of the GUIDE are supported in one of the associated Sourcebooks, which provides case studies of the way in which three countries – the Netherlands, Ireland and Italy have over the years developed their evaluation capacity.

The focus in this part of the GUIDE is on institutional capacity in public administrations. It is of course recognised that evaluation capacities can be conceived of quite differently – as a more dispersed capacity often located in practitioners, communities, among programme managers and in civil society. However, given the concern to strengthen socio-economic programmes at a systemic level, we have chosen to emphasise the administrative point of entry here.

3.1 DEVELOPING INSTITUTIONAL CAPACITY

The nature of institutional capacity

This GUIDE has already painted an ambitious picture of what is expected of evaluation, how it should be organised and what it can deliver. In part 2, for example, we provided guidance to administrations in how they might design and implement an evaluation. This assumed that the administrative and institutional capacity was available. In this part of the GUIDE we are concerned with how to create such capacity so as to make the ambitions of the GUIDE practicable. 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 the evaluation process and its product.

Developing evaluation capacity is necessarily a shared concern of the wider evaluation community, including both those who manage and commission evaluations, those who have an interest in evaluations at a policy and programme level and those who undertake evaluations. Having this capacity adds value to individual evaluation efforts and should be regarded as an integral part of the management arrangements for socio-economic development programmes. It takes time to develop such capacity and the needed structures cannot be put in place once and for all. They need continuous nurturing to deliver sustainable benefits.

'Nurturing' the ability to fully exploit the potential benefits of evaluation in a policy or programme

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 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 a twofold

Building institutional capacity

imperative:

1. To overcome the inefficiencies of the traditional public administrations, by adopting (and adapting) the lessons from the most successful private or non-profit organisations (what is often called the New Public Management movement);
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 or past programmes to reach their objectives in an efficient way, was identified in part 2 of the GUIDE as an important contribution of evaluation. This 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 when designing evaluation questions, 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”. Because contemporary theories of socio-economic development rest heavily on territorially-based (endogenous) resources and potential, such increases in social capital help to promote sustainable socio-economic development.

*Modernisation,
reduced costs
and enhanced
social capital*

Economic models, political systems, managerial doctrines and interpersonal relationships all play a role in the implementation of policies and programmes. Multi-disciplinary contributions need to be brought together in an analytical framework aimed at explaining the variations that can be found in the real world and at exploring the causal links between goals, actions and results. This help to ensure that the kind of knowledge that is generated is useful, and that the costs of the evaluation are not higher than the benefits. This in turn means that the people that are responsible for the policy or the programme must be the first people that must be convinced of the need for evaluation. Whilst their support is essential, they must not be allowed to capture the process.

*Internalised into
decision making*

Optimally, in order to be useful and used, ie in order to make a difference, evaluation must be integrated within regular, decision making and implementation processes. Decision makers at all levels ought to appreciate the usefulness of impartial and independent evaluations that are closely connected to regular management processes. Heavy costs are incurred by forcing an evaluation on unwilling and sceptical decision-makers.

On the supply side, evaluation must be carried out professionally by people that possess the relevant skills and expertise, including the ability to understand how their efforts can meet the needs of the commissioner and programme manager.

Professional strength, is a crucial factor in securing credibility and demonstrating independence and impartiality vis-à-vis the different and sometimes very powerful interests involved in a programme and/or policy. The sources of professional strength and independence are varied. They include:

- Professional norms of behaviour. The evaluator runs the risk of losing credibility through acting in an unprofessional or unethical way;
- Ethical codes that are widely recognised and disseminated. These should support professional norms of behaviour;
- Independent and well-established institutions within which evaluators work can lend their judgements greater weight and allow them to resist external pressures;
- High quality education and training is usually a pre-requisite for professional recognition;
- Professional societies that bring together evaluators with different levels of expertise and experience can be shared and practical problems discussed.

3.2 AN IDEALISED MODEL OF EVALUATION CAPACITY BUILDING POLICY

The key questions of evaluation capacity building are: How can this desirable state of affairs be brought about? Is it possible to overcome the deep resistances, often embedded in the national administrative cultures, to the introduction of evaluation?

Experience shows that it is certainly possible to overcome the kind of resistance that we have described and to embed evaluation in public policy institutions. An evaluation culture has evolved over many years across Europe and, as was noted in part 1 of the GUIDE, has become more widely established in recent years because of the impetus of the European Structural Funds. We are not starting from scratch. As many accounts of the implementation of European cohesion policy show, performing and using evaluation is now an integral part of the path towards better institutional capacity. This takes time and effort but the rewards are considerable in terms of innovative administrative practices more generally as well as in terms of improved programme management.

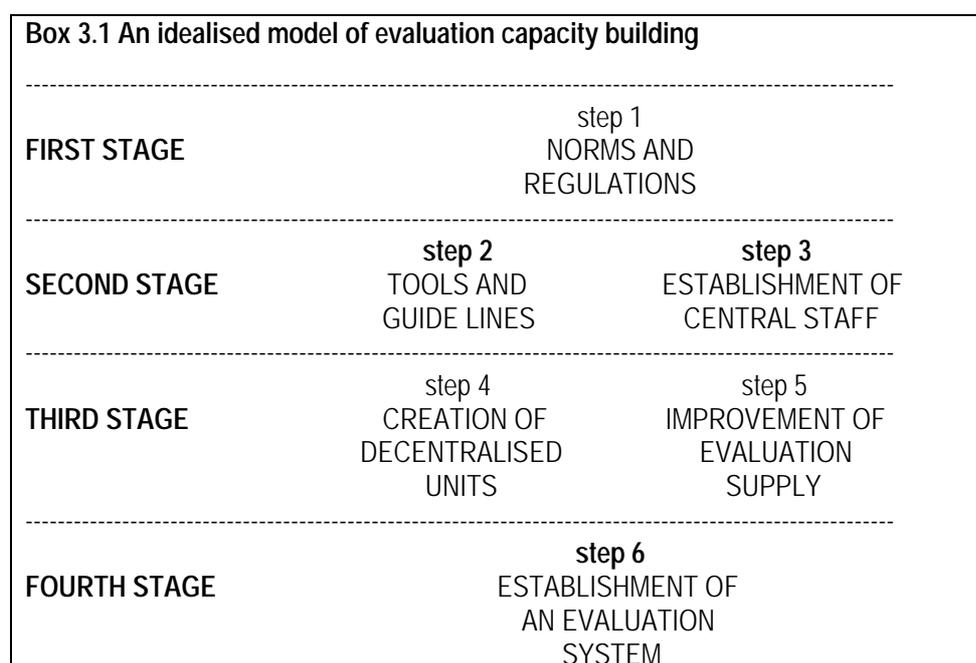
Performing evaluations is becoming more and more frequent

Experience has also shown that there is no one way of introducing these kinds of institutional reforms. This is partly because, in every case, the point of departure will differ. Public service traditions, academic resources, the organisation of professional and technical skills and the nature of the market for selling these skills are radically different in different countries. The development of evaluation capacity has to start with the pre-existing situation and with a diagnosis of actual needs. It should be regarded as a process that will inevitably take time and go through a number of stages. Whilst the first steps may be modest, they

There are different routes to developing capacity

should be taken in the context of a longer-term perspective buttressed by intermediate and final objectives. Such longer-term planning needs to be accompanied by the mobilisation of appropriate resources.

What is presented below is an idealised model: a map for a journey that has a number of stages and intermediate destinations. As with any idealised model, it is only useful to clarify options and exemplify the kinds of choices that need to be made. In practice, countries have followed different routes that do not conform strictly to the sequence presented below.. Nevertheless, as laid out, it represents a plausible and reliable route to develop evaluation capacity. The model can be adapted to diverse circumstances. It is intended to provide guidance both for those who have already developed work in this area as well as those who are starting from a less advanced situation.



Stage 1: Mandating evaluation

Stage 1 in this idealised model, the point of departure of the journey, is very often about accountability. The driving force is, in most cases, an external pressure that requires evaluation through norms, regulations or policy objectives.

This is certainly the case in the context of European cohesion policy. It is indeed the responsibility of the European Commission towards the Member States, the European Council and European Parliament to ensure that the monies employed 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. However, this raises questions about the integrity of the process and its independence. Hence the need for clear evaluation standards and procedures to safeguard the independence and standing of evaluations.

Starting the journey from the needs of accountability and formal regulations

Fortunately, the accession countries have no regulatory evaluation requirement to fulfil with respect to the European Structural Funds before 2006. In the meantime it has been recommended to them that evaluation plans are prepared and budgets are allocated for evaluation activities during the period 2004-2006. This provides a valuable opportunity to prepare the ground for post-2006 requirements. A number of useful approaches could be adopted as 'preparatory activities':

- To involve actual and potential members of central and decentralized units in evaluations to build up their familiarity with the circumstances they will be expected to manage;
- To initiate a small scale support system for self-evaluation that could disseminate a culture of evaluation and raise awareness;
- To invest in new monitoring and data systems that will support future evaluation activity.

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/or open themselves up to some form of external scrutiny.

But relying formal regulations and compliance can be limited

The formal requirement to conduct evaluations is a stimulus for monitoring and evaluation activity.

Box 3.2 Rural Development Programme evaluation, Poland

In Poland the evaluation of the World Bank funded Rural Development Programme concluded that the formal requirement to conduct evaluations played a very important role in introducing monitoring and evaluation in the country. Therefore there is a need to introduce legislative provisions and institutional arrangements in order to stimulate proactive attitudes aiming at systematically applying evaluation also to other programmes.

Putting in place such rules and regulations is the first stage of establishing a policy of evaluation. This will establish an impetus for new evaluation activities and roles. For the first time, documents called Evaluation Reports are circulated and consultants and academics begin to describe themselves as evaluators while administrators are appointed as evaluation managers. The weakness with the formal requirements for evaluation practices is that they are indeed formal and call for some sort

of compliance. Even compliance mechanisms can also be formal and, as such, not usable or used. There are a number of reasons why systems based exclusively on rules and regulations do not work:

- They depend for their implementation on some kind of sanction. 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.
- The establishment *via* formal regulations of the evaluation activities tends to underestimate the supply side, ie the availability of knowledge and skills.

It is once the formal regulations have been established that problems of supply come to the fore. It is because the driving force is external that little attention is paid to quality both methodological and substantive.

As an East-European observer has put it "*many officials having the choice of spending the money on something concrete, like a kilometre of sewage pipe, or something abstract, like an evaluation report, would tend to select the former option*".

External bodies who set up evaluation will often find it difficult to grasp the conditions under which the implementation of the programme or policy is carried out and can therefore be easily misled by those working in the field. 'Information asymmetry' is a term often used in organisational studies to describe these kinds of gaps in understanding between different bodies working together when one is in a formally more powerful position.

For these reasons this first stage of evaluation capacity building, more often than not, is perceived as disappointing and highlights problems of quality, implementation and lack of cooperation between evaluators and evaluatees.

Stage 2: Coordinating evaluation

There are two kinds of actions that those responsible for evaluation can take in response to the limitations of a purely formal and rule-based 'first stage' evaluation policy.

The first is to try to achieve the goal of good quality evaluation through the issuing of guidelines and/or the preparation of tools to be used by commissioners and/or evaluators. This can include the publication of handbooks, the diffusion of standard terms of reference, the building up of sets of indicators, the preparation of ad hoc software, etc. In some cases this takes the form of full guidelines specifying the expected content of the evaluation reports, possibly including the methods to be used. Such an approach is taken in this GUIDE as well as the earlier MEANS Collection.

*Providing
guidelines and
basic tools*

The basic advantage of this approach is the possibility offered to increase the comparability of evaluation studies, making some sort of quality control possible. This is why very often the preparation of guidelines and instruments is considered as a part of a co-ordination effort. The argument stresses the need to avoid having too many different and divergent evaluation approaches.

However, an excessive emphasis on development tools and guidelines is not without risk. The more rigorous the guidance provided, the more likely it is that its enforcement reproduces the shortcomings of the regulatory approach of the previous stage. Following guidelines of this kind can simply reinforce the external character of the whole exercise and reinforce resistance.

Because guidance, tools and standard formats do not tend to work on their own, an alternative approach is to create a central professional staff in charge of evaluation activities.

The emphasis is on the professional quality of the staff. The influence of the staff of such a unit does not derive from the power to intervene in the evaluation process – and even less the policy formulation and/or implementation phase. Rather, these units depend heavily on their knowledge of evaluation and the help they can offer public authorities responsible for the design, implementation and use of evaluation studies.

Professionalising staff as a way of improving quality

The creation of a central staff is probably the most critical step in evaluation capacity building. This can be achieved in quite different ways. The continuum is between total independence and very high status, of which the French *Conseil National de l' Evaluation*, elaborated in Box 3.3, is possibly the most relevant example, to units embedded into central departments as is the case in the Netherlands and the UK.

Box 3.3 Conseil National de l'Evaluation

Since 1990 in France the central responsibility for evaluation has been given to the *Commissariat du Plan* under the supervision of an inter ministerial committee and with the help of a Scientific Council. A special fund was created in order to finance the evaluation studies involving different central departments. After a reform in 1998 the National Committee was created composed of 6 researchers and scholars, 3 representatives of regional and local authorities, 3 representatives of social organisations, 1 member of the Conseil d'Etat and 1 member of the Court of Accounts. The main task of the National Committee is to propose a list of evaluation studies to carry out as well as to secure the scientific quality of them.

There are, in practice, a number of different institutional architectures possible in the way central units are designed. The full range would include:

- A central unit that directly co-ordinates evaluation activities across government departments;

Different pathways may work for different countries

- Free-standing departmentally-based units that are centralised within those departments;
- Inter-departmental networking where a central unit co-ordinates indirectly through linked departmentally-based units;
- Decentralised networked units within departments where key personnel are attached to different policy directorates and come together as a virtual team for co-ordination purposes.

In general the pathway suggested in the idealised model presented in this part of the GUIDE – that begins with a central unit and then ‘cascades’ down through the rest of the administration whilst also taking the lead in developing evaluation supply – appears to fit well in smaller countries. However it is recognised that alternative development pathways are possible. It is certainly true that in some larger countries where there is a strong social science community and where evaluation has developed as a professional practice over many years, a much more bottom up pathway can be observed. However, the main elements of the model described here, have to be put in place for capacity to be institutionalised, whatever the sequence.

Regardless of the institutional arrangements two aspects are very relevant: the mix of professional knowledge and the accumulation of evaluation expertise.

Balancing substantive knowledge and evaluation know-how

With regard to the kinds of knowledge that members of a central unit need, this has to start from the substantive content of the programme or the policy being assessed. It is therefore imperative to strike a balance between the theoretical, methodological and technical expertise in the field of evaluation and the grasp of the substantive policies involved in socio-economic development. Too much emphasis on substantive policy knowledge runs the risk of not paying enough attention to the problems of valid evaluation judgments, and of “reinventing the wheel” as far as evaluation methods and techniques are concerned. Too much emphasis on evaluation experience alone can privilege abstract methodological correctness without tackling substantive questions. This is especially true in the case of socio-economic development policies, where the multifaceted dimension of the problems and of the programmes are accompanied by inherent uncertainties about the ability of the solutions put in place to create the conditions for a self-sustained and sustainable process of development. This requires a pragmatic and flexible approach to evaluation.

It is difficult to advise and influence the evaluations of others if the staff of central units are not themselves experienced in the day-to-day realities of evaluation practice. At this early stage in the development of evaluation capacity, it is unlikely that there are many well-trained professionals who can be recruited to join a central unit. They are more likely to be recruited from neighbouring professions such as consultancy, academia and interested civil servants. There are two approaches to developing necessary expertise. The first is to launch an extensive and tailored training programme. Given the practical nature of the knowledge and know-how required, classroom learning on its own will be of limited relevance. This does not mean that some degree of formal training is not essential, if only to ensure a common approach, minimum standards and a shared language. The second approach of developing

The central team needs hands on experience of evaluation

expertise is through 'learning by doing'. Probably the best and fastest way to build up expertise is to involve the central staff in actual evaluation exercises, sometimes in a joint effort with external experts. This has the advantage of rapidly confronting members of the staff with the difficulties arising both on the side of the managing authorities and of the professional providers of evaluation.

Box 3.4 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.

There is a balance to be struck between central staff involvement in real world evaluations and the need to keep a distance and be respectful of the independence of the evaluators themselves. It is difficult to combine a role of commissioner with the role of executing the evaluation. The staff of the central unit therefore have to manage the tension is between a "hands on" attitude, in which the central staff considers itself the commissioner of all the evaluation exercises, and a "hands off" stance, in which its role is limited to mere stimulus towards quality improvement and towards a larger use of evaluation results. A typical progression is from an initial combined role where parts of a central unit will specify an evaluation and other colleagues will undertake the work, towards a more differentiated practice where the role of central unit staff is largely one of evaluation co-ordination management and stimulus. However, this is not a once-off process of knowledge acquisition. Given the changing circumstances within which evaluations take place and the evolving nature of socio-economic development programmes, this expertise-base will need to be replenished regularly. This is even more important when staff turnover occurs, as is typical in most public agencies. Ensuring that some elements of evaluation tasks are kept in-house will provide learning opportunities for those working in central co-ordination roles.

...but keeping a distance is also important

However, the most important issue that arises during this stage of development of evaluation capacity building is the slow but usually very clear shift of the main purpose of evaluation from an emphasis on accountability to an emerging parallel emphasis on management and learning. Once staff are in post within a central unit, they become concerned with how the outputs of the evaluation system are used. Alongside the development of a professional ethos there is usually a concern for relevance and usefulness. Token compliance with procedures is not sufficient. It will not guarantee that programme managers and policy makers take evaluation seriously. It is at this stage that the importance of cooperation of policy and programme managers with the staff of central units becomes a priority.

Discussions with managers and policy makers will quickly reveal that they will only take evaluation seriously if it is useful to them. Evaluation to be useful must lead to an improvement of the knowledge base for their decisions.

Cooperating with programme managers becomes increasingly important

Working more closely with managers and taking seriously their operational and policy needs shifts the balance of evaluation towards management and learning. Many accounts of the successes of evaluation stress this point: the “discovery” of it by the managers as a tool for learning how to change the ways in which the programmes are built and processed.

In some cases this discovery of the operational benefits of evaluation by managers and policy makers temporarily eclipses the accountability dimension. This is something that should not be feared. Unless evaluations are of good quality and taken seriously because of their usefulness, neither learning nor accountability will occur. And partly this stress on the managerial dimension is a stage in the evolution of evaluation, and is not the final point of the journey.

A shift usually occurs at this stage from accountability to learning

Stage 3: Institutionalising evaluation

Once a central unit is up and running, a third stage of evaluation capacity building begins: that of institutionalisation. This is composed of two different steps, usually taking place more or less simultaneously.

The first step is to strengthen the demand side of evaluations through the creation, within the administrations that manage programmes, of a network of specialised evaluation units. This is most likely to happen in larger countries and usually starts from the departments or agencies that are mainly involved with the implementation of socio-economic development programmes.

Once a network of evaluation units is established so is the beginnings of an evaluation ‘community of practice’

Particular attention is to be paid to the staffing of those units. If the involvement of junior staff in evaluation work is usually an excellent method of providing training, equal attention has to be put on ensuring that these units are directed by people senior enough to have credibility with decision makers and evaluators. This will help reduce excessive staff turnover.

The establishment of a plurality of specialised units has two consequences. First there will be the beginnings of an emerging ‘community of practice’ ie a significant group of staff with a shared professional interest working within comparable settings. Second there will be an extension in the scope and scale of evaluation within the relevant public sector.

The staff of a central evaluation unit may be located in different parts of an administration. The most likely location is as part of the planning and programming department. As such it will share the typical concerns of such a body. This will give it an initial emphasis on planning and feasibility at the ex-ante stage of the evaluation. This will involve analysing the initial planning documents in order to assess their coherence, intervention logics and evaluability. The emphasis on ex-post evaluation – assessing the results and impacts of the programmes

– is likely to be less at the beginning. The implementation concerns typical of mid-term or ongoing evaluations do not usually take priority at this stage. The usual temptation of a central body is to emphasise the need to streamline and develop standard procedures for the implementation of the programmes and projects and to ensure that the implementing authorities comply.

With the diffusion of the evaluation capability down the implementation line, enhanced by the creation of evaluation units within or near the managing authorities, a subtle change of focus occurs. Initially the concerns are mostly about the ability to carry out the complex programmes designed in the planning phase within the rules and regulations constraining the public administration. Consider for instance the difficult situation in which the people in charge of the regional programming documents in European Structural Funds find themselves. They must try to spur socio-economic development in their territories, a difficult enough task in itself, whilst at the same time:

- Streamlining the transversal priorities – from equal opportunities to information society – that are mandated;
- Coping with the complex procedural arrangements that the European or national regulations prescribe (e.g. the transparency requirements of financial control);
- Paying attention to the automatic decommissioning of the so-called n+2 rule (the prescription that money must be spent within two financial years from the moment in which it is allocated).

It is not surprising therefore that line managers are sometimes resistant to implementing the evaluation activities mandated by, say, European cohesion policies. They may consider these requirements as an additional procedural impediment to the “smooth” running of their programmes. But this is also the reason why the establishment of decentralised evaluation units may represent a major step not only in evaluation capacity building, but also for the whole socio-economic development policy. The creation of decentralised units closer to programme managers (eg within their departments or agencies), makes it easier for accommodation to be reached with evaluation priorities.

The perception of evaluation as an additional burden for busy managers is common.

This is one of the reasons that central units refocus on management concerns

This can also be achieved through the use of financial incentives. Box 3.5 provides an illustrative example.

Box 3.5 Use of Performance Reserve in Italy

In Italy in the use of Structural Funds the performance reserve in Objective 1 programming documents was supplemented by 6% in order to incorporate several measures related to institutional capacity building. In particular the timely awarding of the contract to the mid-term independent evaluator and the setting up of the decentralised evaluation unit were both considered as relevant elements and on this specific point the compliance was generalised. Attaching a monetary value to the various elements the national and regional administrations were stimulated to act in order to reach the re-programming stage with the evaluation institutional framework well in operation.

As central units become established they begin to refocus evaluation activity. This would be so even at the ex-ante phase when evaluation becomes used to choose between the different projects to fund,

transforming it to become part of the decision support system. Even more important is the emergence of mid-term evaluation as a crucial element of programme implementation. Its main purpose is to foster a learning culture throughout the implementation chain. It becomes a means to manage the various constraints set by programme time-scales with the more ambitious goals of institutional development. This is also where formative evaluations become critical. Mid-term evaluations identify necessary managerial adjustments and the kinds of corrective action that need to be taken by wider partnerships and stakeholders. This more direct involvement in management decision making is again eased by the creation of decentralised units within the operating and policy departments of public authorities.

This extension of the scope of evaluation creates a further extension of the evaluation supply chain to involve those who supply evaluation services as well as those who commission evaluation. At the earlier stages of the development of evaluation capacity, there is as we have seen, reliance on tools and guidelines and the competence of central unit staff to manage evaluation supply. This becomes more difficult as the system expands and more so when specialised knowledge is required within substantive areas of policy and within particular areas of evaluation methodology.

Once the focus shifts from ex-ante evaluations that may well be conducted in-house, as we have seen especially in the early stages of capacity development, the role of independent suppliers increases. This is especially true in the case of the European Structural Funds, which requires the independence of the suppliers for the mid-term evaluation. This makes the evaluation demand strongly dependent on the characteristics of the market.

Improving the supply of evaluation expertise usually starts to involve evaluation education initiatives

This is why in an ideal evaluation capacity building strategy, needs to initiate activities to improve evaluation supply in parallel with the establishment of decentralised evaluation units.

Box 3.6 Ex-post evaluation of Phare

The ex-post evaluation of the Phare support 1997-1998 was conducted through an original 'learning by doing' method in which the European Commission who was in charge of the entire exercise involved the national authorities in drafting the terms of reference, recruiting the evaluator teams and supervising the work through national representatives with little previous experience of evaluation. The 25 external evaluators hired through a selection process had a very varied professional capacity in evaluation ranging from very strong to very weak. Both the national representatives and the external evaluators were involved in a learning exercise that involved workshops, web based e learning and joint work with EU consultants.

There is not a single way of improving evaluation supply. Three approaches are common:

- To build up relationships with educational institutions, in particular, universities;
- To support the development of a professional evaluation

community;

- To develop transnational exchange of evaluation experience.

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 excellent able to undertake research and 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 eg theories and methods of socio-economic evaluation into the formal curricula of economists, engineers, sociologists, planners, etc. It is important that some specific issues, relevant to the field of socio-economic development (such as transversal priorities like information society, equal opportunities and the like), are addressed as part of these programmes

Working with Universities

It will be important when planning education and training programmes to recognise the practitioner and professional aspects of evaluation work in the socio-economic domain. What needs to be developed are practitioner skills and know-how as well as academic knowledge. This can be achieved by including within curricula less formalised pedagogic approaches (eg workshops, training seminars, guest lectures from practitioners, study visits to administrative bodies with responsibility for socio-economic development etc.).

Practitioner as well as academic knowledge is needed

The development of a professional community encompasses both the supply and demand sides of evaluation practice: they need to involve both the commissioners of evaluation ie those working in central and decentralised evaluation units and the providers of evaluation. The main vehicle for such development is professional networks such as national evaluation associations. These typically bring together people from different institutions (academia, public administration, research centres and consultancies), with different disciplinary expertise (sociologists, economists, planners, political scientists, 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. Box 3.7 provides links to 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 Part 2 of this GUIDE, 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.

Professional networks and evaluation societies involve evaluators and commissioners

Box 3.7 URLs of European Evaluation 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/>
Walloon Evaluation Society: <http://www.prospeval.org/>
Netherlands: <http://www.videnet.nl> and <http://www.beleidsonderzoek.nl>

There has been a general trend towards internationalising evaluation practice in recent years involving exchanges of evaluation experience among administrations and between practitioners. This has partly been at the initiative of international bodies such as OECD, the World Bank and the European Commission that have sponsored a series of international conferences on evaluation, but has also occurred on a bilateral basis between different countries wishing to learn from each others' experience. Many of these bilateral initiatives have involved members of national evaluation societies and societies themselves. Most recently a number of international evaluation bodies have been created such as IDEAS (International Development Evaluation Association) and IOCE (International Organisation for Co-operation in Evaluation).

One of the rationales for such international exchange is the limited range of evaluation in any one country. Furthermore when evaluation activities have the European dimension, there is an additional argument for drawing on pan-European experience. There is also a linguistic element which cannot be ignored. Like all applied social science activities, evaluation reports are written in the native language of evaluators. As a result the readership of evaluation reports is sometimes very small and evaluators have very limited knowledge of actual examples of evaluations performed in other countries. Languages that are better known such as English and French are more likely to be internationally accessible. But this means that most evaluation work undertaken in the Northern, Eastern and Mediterranean countries becomes lost as far as good practices are concerned. The development of transnational exchange of experiences, therefore, is an important vehicle for disseminating such experience and thereby improving evaluation supply. However, even international efforts at exchange can be undermined by linguistic factors. It is usually the case that those who are active in these kinds of exchanges are French or English speakers, continuing the marginalisation of other languages and limiting those from different countries able to participate actively in debates.

*Internationalising
evaluation*

Stage 4: Towards an evaluation system

The final stage of evaluation capacity building is the creation of a fully-fledged evaluation system in which evaluation is incorporated into policy making, programme management and governance. This stage consists of:

*Creating new
horizontal
networks*

- The establishment of stronger internal links within the system
- Opening up the network to external stakeholders.

Until now, in our idealised journey, the logic of the policy has been mainly top-down. Following this logic, the links within the system have also been vertical rather than horizontal. Once the different units are in place, it is possible to develop feedback mechanisms from the periphery to the centre. For example, patterns of communication will be established between different regions and territories, perhaps by the creation of practitioner networks, common training programmes and a professionalised cadre of sectoral evaluators. This feature changes the form and workings of an evaluation system and is crucial for the full realisation of learning potential that is the goal of the entire exercise.

At earlier stages in the development of evaluation capacity, evaluators have inevitably interacted with policymakers and programme managers but probably not with the wider constituency of stakeholders. Once an evaluation system becomes mature it begins to involve a range of institutional stakeholders that represent different groups implicated in socio-economic development, for example, regional governments, parliaments, municipalities, public and semi-public agencies etc. There is also a commitment to involve wider civil society groupings that bring policymaking into touch with the citizens, for example local associations, NGOs and the media.

Broadening the base of stakeholders and partners

As an evaluation system becomes mature and more complex, ensuring its coherence becomes more of a challenge. Whereas in earlier stages the emphasis might well be on technical skills, at this stage, there is a premium attached to communication skills within the evaluation community. There is also a need to ensure that actors present within the wider evaluation system have some clarity about their respective roles, for example, what are the main concerns of central government agencies vis-à-vis local and regional governments? And what are the training priorities of different actors and how far do they share the results of their investments in training and professional development?

The re-emergence of accountability-with greater reliability

Box 3.8 Evaluation culture

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.

At this stage of evaluation capacity building the general evaluation purpose of accountability, that was pre-eminent at earlier stages but subsequently was overshadowed by the imperatives of learning again comes to the fore. However, there are subtle and not so subtle differences. Accountability can at this stage encompass systematic reports based on research evidence and practitioner experience that is far more reliable than the kinds of accountability reports produced largely for compliance reasons at an earlier stage. The system has

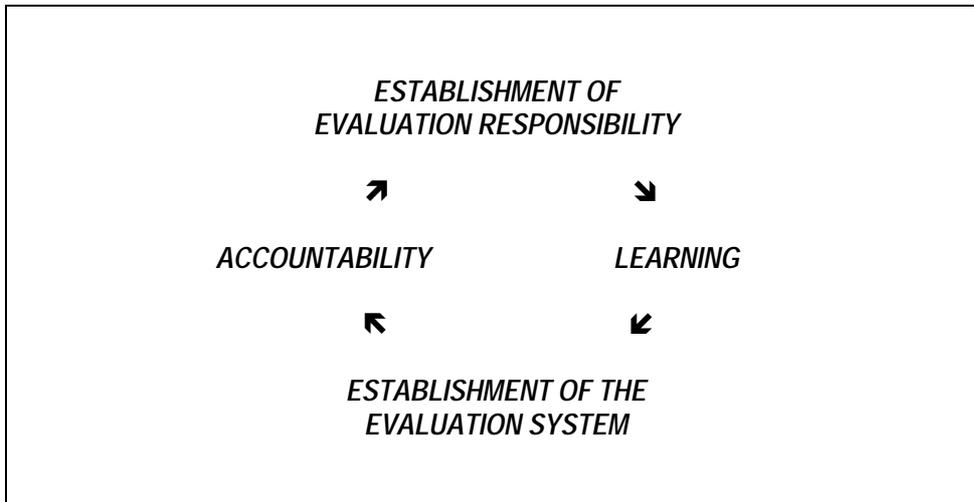
become capable of producing more reliable accounts of its performance because it has learned how to do so at a systemic level. This is why evaluation experts continue to emphasise the centrality of learning as the following slogan suggests:

'instead of putting our efforts into learning to become accountable, it is better to stress the fact that we are accountable for our learning'.

This is much more than a catch phrase, because it emphasises a main responsibility of politicians and professionals alike in many contemporary complex and uncertain policies, and certainly in socio-economic development. This is to recognise and correct mistakes, to generalise good practices, and to (try to) steer and orchestrate a whole set of different actors each with their own interests, preferences and resources, rather than simply fulfil the promises laid down in some planning document.

Box 3.9 depicts the ideal journey that we have just described.

Box 3.9 The virtuous circle of evaluation



3.3 STRATEGIES FOR DEVELOPING EVALUATION CAPACITY

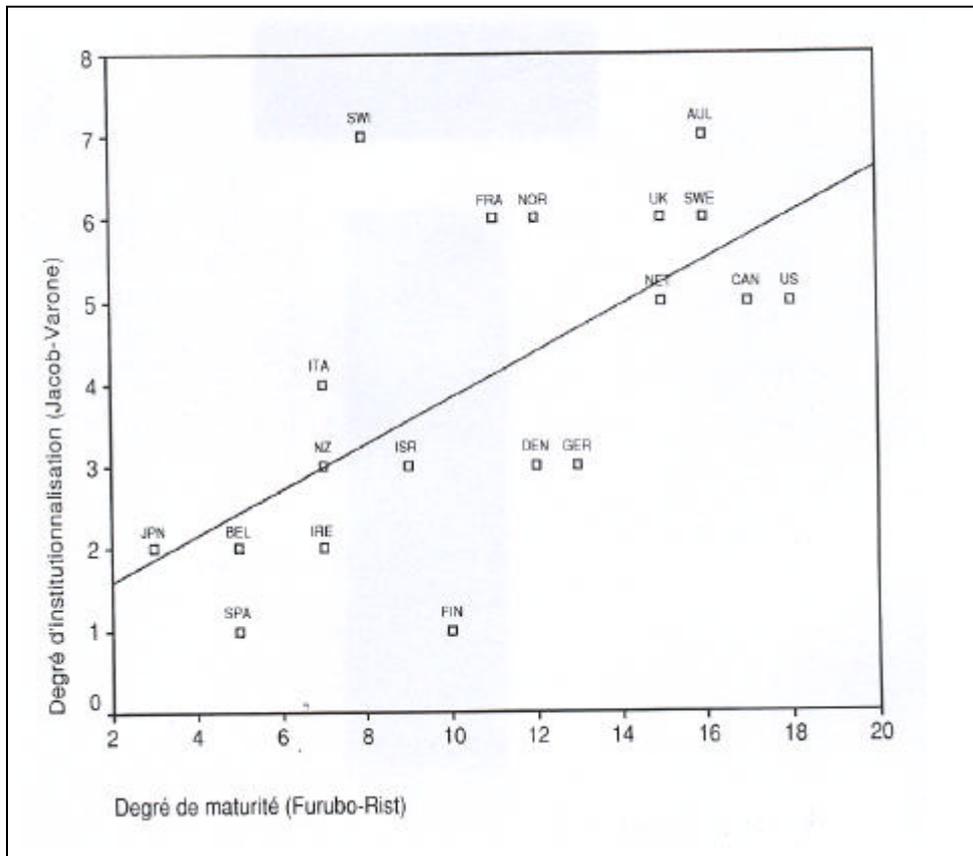
In a comparative perspective it is likely that no single industrialised country has reached the final stage of evaluation capacity development, but they are still struggling between the second and the third stage depicted above. Box 3.10, drawn from recent work, shows the level of institutionalisation and maturity of evaluation practices in 19 countries.

Of course, this model as with all ideal models, has to be treated with caution when devising an evaluation capacity building strategy and policy. For example, in many countries the setting up of decentralised evaluation units preceded rather than followed the birth of the central staff. In the UK for example, there are strong departmentally based evaluation units but no central unit that co-ordinates these departmental activities. Furthermore the professionalisation of evaluation and the growth of evaluation societies has often come about spontaneously rather than part of any conscious public policy.

Developing evaluation capacity is part of a wider agenda of public sector modernisation

Incidentally one must note that building evaluation capacity – defined as the ability to commission, carry out and use evaluation studies – is but a small part of a larger agenda of public sector modernisation and adapting the State to the new demands of contemporary societies. Socio-economic development policies require both these wider reforms as well as the development of evaluation capacity.

Box 3.10 Institutionalisation and maturity of evaluation in 19 countries



3.4 GOLDEN RULES

1. Evaluation capacity will not develop on its own. It needs to be planned and it will take time. Four to six years may be required. There needs to be a policy, that is a set of priorities and a strategy that, over time, increases capacity both in relation to the demand side (those who commission evaluation) and the supply side (those who provide evaluation services).
2. An underlying principle that needs to be built into any evaluation system is the professional independence and impartiality of evaluation, given the vested interests that are likely to seek to influence evaluation and its outcomes.
3. At the early stages of developing evaluation capacity there needs to be some element of formal regulation to support the obligation to evaluate. This creates a legal framework within which subsequent organisational innovations can occur and provides a means of overcoming the initial resistance that will undoubtedly be

encountered.

4. Although there is no single strategy for developing evaluation capacity, there are certain elements such as the creation of specialised units within public administrations and the initiation of training and professional development initiatives that are central at an early stage.
5. The trade-off between different evaluation purposes eg accountability, management and learning has to be contrived - and the balance will change at different stages in the development process. It is not sensible to regard these trade-offs as problematic, they just need to be managed.
6. As has been emphasised elsewhere in this GUIDE we need to keep in mind that we are essentially concerned with the evaluation of socio-economic development. Many aspects of this strategy that has been outlined above, in particular, the importance of vertical and horizontal links between different tiers of government and progression towards involving more stakeholders right the way through to civil society follows from the requirements of socio-economic development.
7. Although this GUIDE takes an essentially top-down view emphasising the responsibilities of public authorities and governments, this is not the only way in which evaluation capacity can be developed. In some larger countries with a strong social science research tradition and an active civil society, it has been known for evaluation capacity to grow up substantially outside the ambit of the State. This emphasises the responsibilities of all parties when developing an evaluation system and an evaluation culture. Even if the initial responsibility rests with public authorities, it is desirable to move rapidly to the point where responsibilities are taken on board and shared by professional groups and, indeed, citizens wishing to secure their democratic rights.
8. Although much can be gained from an initial and planned investment in developing evaluation capacity, this is not a one-off process. As new programmes are launched, new agencies created and new staff recruited, many of the same processes of building links, training staff, clarifying roles and supporting independence and professionalisation will need to be revisited, otherwise well developed systems can easily fall behind, resting on their laurels and failing to adapt to new circumstances.
9. The supply and demand side of evaluation are interdependent. Not only should efforts be made to avoid a gulf developing between those who commission and those who deliver evaluations, but co-operative links need to become the norm. This can be supported by the development of professional communities, evaluation societies and education programmes in which both sides of the evaluation 'business' are involved.
10. Reference has been made to evaluation standards, guidelines and ethics in this part of the GUIDE as in previous parts. Experience in

the world of standards has shown that top-down initiatives to impose standards are rarely effective. Such formalised systems of standards are best developed after a period of exchange and debate within a community of practice. The term 'de facto' standards has been used to describe the process by which the different parties: practitioners, academics, civil servants, grass-roots managers etc come to share sufficient agreement for more formal standards to be created.

11. We have noted that the re-emergence of accountability in a more sophisticated form at Stage 4 of the evolution of the idealised model of evaluation capacity development that has been described. Implicitly we have begun to touch on issues of performance management which are also in an integral part of contemporary public sector reforms and the new public management. Experience would suggest that it is only at Stage 4 that systems become robust enough to incorporate performance management alongside other purposes of evaluation without these being undermined. This does not mean that one should ignore the differences between evaluation and performance management anymore than one should ignore the differences between evaluation and audit. However, a robust evaluation system should be able to provide relevant information also for performance management purposes.

PART 4 CHOOSING METHODS, TECHNIQUES AND INDICATORS AND USING EVIDENCE IN EVALUATION

This part of the GUIDE 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. In part 2 of the GUIDE, 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 raise 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. The first section of this part of the GUIDE considers: 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.

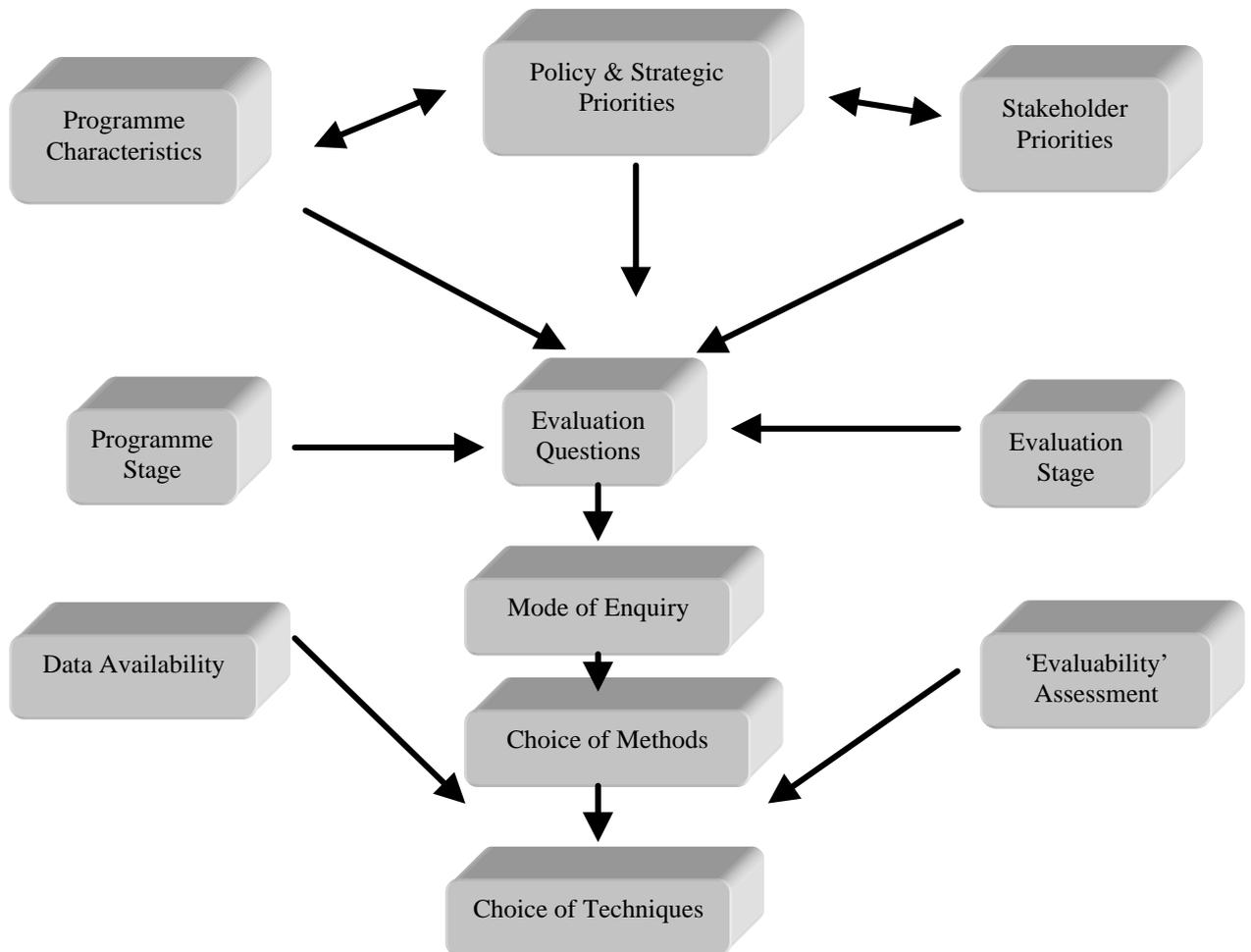
4.1 FACTORS INFLUENCING THE CHOICE OF METHOD, TECHNIQUES, DATA AND EVIDENCE

Choosing methods and techniques

As been discussed in part 2 and 3 of this GUIDE, 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. Box 4.1 positions the choice of methods and techniques within this broader context.

The choice of techniques depends upon purpose, object, timing and data availability.

Box 4.1 Choosing Methods in a Wider Context



This part of the GUIDE fills out many of the cells in the above figure as this is the only way to ensure that the choice of methods and techniques take into account all the relevant factors.

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 Part 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 for. Nonetheless, we assume in this part of the GUIDE that evaluators will have scope to choose appropriate methods and techniques, and that the commissioners

Methods and techniques follow from evaluation questions and design

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 Part 1 that the characteristics of an evaluation object, in this case some kind of socio-economic development intervention, are important determinants of evaluation design. (This factor is considered further in Section 4.2).
- *Type of evaluation purpose.* As noted in Part 1, evaluations can have quite different purposes ranging from accountability through to improving management and explaining what works in what circumstances. These different purposes associated with different forms of enquiry also have implications for the methods and techniques chosen. (This factor is considered further in Section 4.3).
- *Different programme/policy stages.* 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. (This factor is considered further in Section 4.4).
- *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 factor is considered further in Section 4.5).

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 work, 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').

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: '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. (In terms of data this argument is elaborated further in Section 4.6).

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.

The distinction between quantitative and qualitative can easily be blurred

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 outcomes 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 over time 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?

Statistics need to be matched to the diverse contexts of socio-economic development

In summary the strength of quantitative evaluation is that it can:

What quantitative methods can - and cannot do

Allow aggregate judgements to be made. Policy makers want to know in

the round whether a policy or programme is working. Aggregate results – have more people got jobs across say 100 programme sites? - will provide material that will support such judgements. However, aggregate measurements will not 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 noted 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 is available (see Section 4.6 for definition of ‘ratio’ data) 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 part of the GUIDE 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’ – what they can do.

Qualitative methods 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.* This is 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 judgement of stakeholders whose intervention logics and programme theories evaluators want to elicit. Or they may be the judgements and experiences of the intended beneficiaries of socio-economic development.
- *We are interested in ‘bottom-up’ understandings.* These can include: the development ambitions of grass-roots actors (small firms, municipal authorities, professional associations) and the expectations and experiences of local people in a local development setting. Such bottom-up information is difficult to fit into top-down categories – highly varied and qualitative understandings are needed.
- *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.* As was noted in Part 2, 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.

Why qualitative methods matter in socio-economic development evaluation

Obtaining and using data and evidence

As was clear in Part 1 of the GUIDE, 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 (eg

Being pragmatic about data –

positivists) will have a different view of evidence and data than those who are more concerned with the way perceptions and theory influence observations (eg constructivists). In this GUIDE 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 judgments 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.

whatever our philosophies are

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.

In this part of the GUIDE we also discuss various issues concerned with evidence and data. In particular, evaluators use very different types of data. Some data pre-exists an evaluation and will come from administrative sources (eg 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 in Part 2, putting in place basic data systems should be part of the planning of programmes as well as evaluations.

Paying attention to data sources improves evaluation quality

4.2 METHODS AND TECHNIQUES FOR EVALUATING DIFFERENT TYPES OF SOCIO-ECONOMIC INTERVENTIONS

In Part 1 of the GUIDE we noted some of the general characteristics of socio-economic development programmes as well as the EU policy context. At a European level there are particular types of socio-economic intervention, which have implications for the choice of methods and techniques. This is not to suggest that there is a one-to-one correlation between any single technique and any particular type of intervention. Nonetheless, certain broad associations of methods with types of intervention are common.

Some of the main types of interventions are:

- Thematic interventions;
- Policy / Sectoral priorities;
- Local and territorial development.

Evaluation of thematic priorities

A specific type of intervention that is characteristic of socio-economic development and features prominently in European Structural Funds, is described as a 'thematic' priority. This occurs where there is an overarching strategy, which is then expected to be inserted or more formally 'mainstreamed' within a range of interventions.

Evaluations of 'themes' require the aggregation of evidence from across many specific interventions to understand the ways in which strategic priorities are being pursued. This can involve the construction of indices that are then added-up across several settings. However usually it is difficult to find meaningful indicators that will work in different settings. Such thematic evaluations are therefore likely to look for different measures in different settings and then qualitatively assess what they add up to. Qualitative assessments of how far thematic strategies conform with policy criteria are also useful: including for example surveys of beneficiaries, interviews of stakeholders and case studies of observed change.

Simply adding up thematic results which are embedded in many programmes is difficult

Examples of such thematic priorities within current Structural Funds settings includes:

- *Equal opportunities.* This is usually conceived of in terms of equal opportunities between women and men but increasingly is applied to other areas of discrimination or inequality such as those that affect people with disabilities or other minority groups including ethnic communities and refugees. Methods appropriate for this thematic area are usually different at a macro and micro level. Take for example the employment circumstances of women. At a macro level, quantitative techniques will normally be used to track women's progress in the labour market, for example, in terms of participation (percentages of women active in the labour market), wage levels, unemployment levels etc. Such methods are likely to include postal surveys, the analysis of administrative data held by employment offices and the reanalysis of available employer-survey data on wage rates. Certain aspects of labour market experience will, however, need to be approached in qualitative ways. For example, when considering working conditions, interviews with people in employment are a necessary complement to more aggregate survey methods. At a micro level, examining the experience of those who are discriminated against in the labour market or how work and family life is (or is not in balance), will require qualitative methods such as interviews, case studies, focus groups etc. This will be the only way in which the judgements of those who have experienced discrimination can be adequately described.
- *Increasing institutional and administrative capacity.* As we have noted in Part 3 of the GUIDE, institutional and administrative capacity is an

important aspect, not only of evaluation capacity, but also of the way in which the public sector and the State responds to new demands for accountability and transparency. Within socio-economic development, we have already noted how partnerships are a pervasive feature of the way in which such programmes are delivered. Evaluating the way in which institutions and administrations adapt and the efficacy of partnership arrangements is therefore another thematic priority. Whilst the outputs of different institutions and administrations can be monitored in terms of indicators of performance and efficiency, these only make sense when they are contextualised in detailed descriptions of organisational arrangements. This requires the use of descriptive field methods such as case studies, observational techniques, interviews, documentary analysis (such as the analysis of the minutes of meeting or organigrams).

- *Sustainable development.* This is typical of a theme that cuts across virtually every policy area, involves many interacting factors and requires, therefore, multiple methods and techniques in any evaluation. This is likely to involve economic techniques such as cost benefit analysis or model resource allocations and their consequences. The recent thematic evaluation of Sustainable development and the Structural Funds used a large number of case studies to test the Four Capitals model of sustainable development and explore its implications as a planning and analysis technique. A strong onus was put on the expert opinion of the evaluators who had both strong backgrounds in aspects of sustainable development as well as appropriate evaluation skills. Given the prospective focus in sustainable development on the consequences for 'future generations', prospective methods such as Delphi (a technique for consulting experts about their judgements) or Foresight methods are especially appropriate. Previously we identified the 'Prospective' category as one important class or 'mode of enquiry' in research and evaluation studies. Sustainable development is a good example of where this mode of enquiry is relevant. However, the sheer scope of what can be encompassed within sustainable development raises familiar problems of aggregation and synthesis. How is it possible to find a composite descriptor or indicator that will sum up what has been achieved across many different interventions? This is a common dilemma in other broadly-based thematic areas.
- *Promoting social inclusion.* As was noted in Part 1 of this GUIDE reducing disparities across the EU has been a consistent theme of EU policy. This was further strengthened following the Lisbon Summit and is well exemplified in the European Employment Strategy, which emphasises social inclusion and 'greater social cohesion'. Broad sets of methods such as target setting and particular techniques such as indicators are commonly used to measure and steer social cohesion. These are used in the National Action Plans on Social Inclusion and the associated reports which Member States have been preparing since 2001. Here also there is scope for more qualitative and participative methods. As was noted with regard to equal opportunities, the judgement of those who are the intended beneficiaries of interventions to prevent social exclusion, is a key element in their success. Interviews, surveys, focus groups, case studies and observation techniques will all be relevant at this more micro level.

- *Information society.* Again following Lisbon and reinforced by the Stockholm Summit, developing an information society and a knowledge-based economy, is seen as a key issue both in terms of a European competitiveness and social cohesion. This is therefore a theme that is being pursued in many aspects of socio-economic development in Europe today. As is common in areas where European competition is at issue, benchmarking with other parts of the world, especially North America and Japan, is a common technique. Where particular interventions are planned, such as those included within the eEurope initiatives, explanatory methods that tried to show before and after effects of interventions would be common. However, the scope of the interventions that are possible under the information society banner are exceptionally broad. They include interventions at the level of regional infrastructure, education, telemedicine, and access to technology for disadvantaged groups. This therefore poses not atypical problems of aggregation and synthesis. The challenge here, as in other thematic areas, is to find ways of adding up a host of different interventions to inform a coherent judgement of what has been achieved. This would often be attempted through composite indicators although open to the criticism of adding up apples and pears. Alternatively, evaluators may segment the theme into particular areas of intervention, measure what has been achieved in these particular segments and then offer a qualitative judgement as to overall progress.

Policy and Sectoral priorities

It was noted in Part 1 of the GUIDE that 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 come to the fore. There is an extent, therefore, to which particular methods and techniques are properly associated with particular policy areas. For example:

Evaluations methods need to match policy areas being evaluated

- 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 criminal justice and human rights interventions (as foreseen in various policies within the Commission's post-Tampere 'justice, freedom and security' agenda) will tend to track changes in outcomes over time (eg in migration and asylum) by analysing administrative data and conducting surveys of citizens both in the EU and in third countries (eg the potential victims of human trafficking).
- Evaluation of active labour market policies and education and training

interventions make use of beneficiary surveys and panel or cohorts 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.

- 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.

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 in Part 1 of this GUIDE. There is a tendency for evaluators who work intensively in particular policy areas such as - environment, human resources, science and technology, justice and crime – 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 tend not to be familiar.

Important that evaluators do not become too narrowly based in one policy area methods

Local and territorial development

Socio-economic development has always had a strong territorial dimension. Over recent years local development (influenced, as we have seen in Part 1 of the GUIDE, by theories arising from the New Economic Geography) 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, 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) can constitute the majority of inputs.

Econometric models are appropriate where the programmes covering 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.
- 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.

Econometric techniques apply when interventions are large

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

Shift share and input output

include:

may be applicable.

- 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 and 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.
- Comparison 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.

Interventions of defined scope can make use of quantitative analysis techniques

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

participatory methods match the bottom-up design of local development programmes

sometimes favoured (e.g. benchmarking), requires that standard indices and measures are applied in order to judge outputs and outcomes.

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 Approaches).

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. Whilst this can sometimes be achieved using indicators, there is also a need for developing models that show the relationship between different factors and different interventions.

4.3 METHODS AND TECHNIQUES 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 in Part 1 of this GUIDE. The purposes 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 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 difficult to apply except in very unusual circumstances. It is 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.

Reliable knowledge comes from accumulating across individual evaluations

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.

4.4 METHODS AND TECHNIQUES APPLICABLE AT DIFFERENT PROGRAMMES/POLICY STAGES

The importance of the time-cycle in programmes and policies has been a theme throughout this GUIDE. In European Structural Funds this is formalised in terms of ex-ante, mid-term 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 outcomes and impacts for intended beneficiaries or territories in relation to intentions (eg 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.

Both economic and statistical and qualitative and participatory methods

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.

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 rather than 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. For example, improved professional development in managers may lead to them leaving an area rather than taking employment locally, thus undermining an objective to develop human resources for local SMEs.

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.

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 innovation 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.

Formative evaluations can track innovative programmes over time

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.

Conclusions / Results: Outcomes and impacts

Policy makers for accountability reasons and key stakeholders, because of their own needs and commitments look to evaluation to provide information on outcomes 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.

4.5 METHODS AND TECHNIQUES APPLICABLE TO DIFFERENT STAGES IN THE EVALUATION PROCESS

Evaluations follow naturally through a number of stages. These usually include:

- Scoping and structuring evaluation work
- Obtaining and analysing information
- Informing evaluative judgements
- Communicating evaluation findings

Sourcebook 2 presents the methods and techniques within these categories.

Scoping and structuring

At the early stage of an evaluation there will be a need to scope and structure the work. Evaluability assessments conducted at this stage can, for example, help structure or further define evaluation questions and identify what data will be needed to answer these questions. As most evaluation depends on criteria, clarifying these criteria is also occurs at the structuring stage. What, for example, would be an acceptable criterion for reducing social exclusion and what level of increased market share is seen as desirable for local firms expected to become more competitive?

Evaluability assessments will help structuring

This is likely to be the stage within the evaluation when indicator systems

will be set up (this is elaborated in greater detail below). However, the process of scoping useful indicators will also identify what kinds of data and evidence are available. The kinds of methods and techniques that are useful at this stage will derive more from project planning than from evaluation per se. However, there will be some methods and techniques such as case studies, network analysis, stakeholder consultation, concept or issue mapping that will also be relevant. The purpose of any of these techniques is to ensure that evaluation resources are used well to decide where they should be concentrated. For example, there may be many interventions and many beneficiaries. The decision at this stage is to decide which should be the focus of attention in the main evaluation phase. This will also be the stage at which specific tools are designed, for example, questionnaires, surveys, and protocols for statistical analysis.

Obtaining and analysing information

Specifying where relevant information will come from is an essential part of scoping activities. However, this is grouped here together with analysing information because analysis and data collection need to be seen as closely connected. The range of analytical techniques is as varied as those that evaluators may use. For example:

- *Delphi surveys* may be used to estimate impacts by involving panels of experts to apply their judgements to available data;
- *Statistical models* can be used to estimate impacts drawing on different types of administrative and macro-economic data;
- *Interview material* can be analysed using computer programmes that are designed for textual analysis;
- *Regression analysis* can be used to estimate the significance and strength of relationships between interventions and observed effects;
- *Content analysis* can be applied to different kinds of qualitative material, including that obtained from interviews, participant observation and case studies;
- *Factor analysis* can be used analyse questionnaire data in order to identify underlying patterns and typologies.

In general, it is a good principle in the gathering and analysis of information to ensure that data collected is as close as possible to the source of the phenomena concerned. For example, direct observations and immediate records of expenditure are always preferable to aggregated proxy measures. Similarly, analysis should not interfere too fundamentally in the content or form of raw data collected. The implications of these principles in terms of analysis are that straightforward analyses are generally preferable to more technically sophisticated analyses. A particular example in terms of survey analysis would be the advantages of drawing conclusions from the analysis of variables for which direct data is available compared with drawing conclusions on the basis of factors that rely on several combinations or manipulations of that data.

Informing evaluative judgements

Making judgements is an important part of evaluation activity. This can be made easier in the narrow circumstances where standards or criteria have been specified in advance. However, in complex programmes as would

be typical of socio-economic development interventions, judgements have to be made when criteria are not clear cut or when data available is open to different interpretations.

It should be emphasised that in most cases, judgement is not a technical matter. It consists of weighing up evidence, applying judgement based on experience or tacit knowledge that is difficult to explicate. Furthermore, judgements in evaluation are usually value-based. They require the application of values rather than the application of techniques. However, this opens up opportunities for certain kinds of technical support for evaluative judgements. For example:

Making judgements is not just about techniques – even though they can help

- *Expert panels* can be used to make a synthetic judgement across different parts of the programme where it is difficult to reach a consensus from the information available;
- *Multi-criteria analysis* can be applied to projects where there are different criteria that might be applied when forming a judgement;
- *Impact assessments* of various kinds (eg environmental impact assessment, gender impact assessment etc) can be used at the results point in a programme evaluation as well as at the design stage.

Other tools such as cost-benefit or cost-effectiveness analysis and statistical models can be used in specific circumstances. However in general these models will provide inputs into judgements rather than judgements of themselves. Someone will still have to attach a value to the information that such techniques will produce.

Communicating evaluation findings

Techniques and methods for communicating evaluation findings are more akin to following professional rules of thumb than techniques per se. For example, with regard to producing reports, it will improve the communication of findings if:

- There is a clear executive summary not exceeding 10 pages in length;
- Reports are kept short and concentrate on the main conclusions and findings whilst additional material is annexed;
- Care is taken to ensure that the supporting evidence for conclusions and recommendations are included and clearly signposted;
- The style and vocabulary of the report is accessible to a lay audience including policy makers and programme managers;
- Where technical terms and abbreviations are necessary, these are defined in a glossary;
- A brief summary is included of the methods used and why these were chosen;
- The policy context is presented in order to ensure that policy makers can see the relevance of the evidence provided.

Communications is about tacit knowledge, rules of thumb and experience

There are a range of techniques available to represent complex quantitative data in an accessible way. Classically these would include tables or graphs. Multi-dimensional 'diamonds' or spider diagrams as they are sometimes called, can be one way of representing such data to a lay audience. Information technology also permits new forms of graphical representation including GIS (Geographical Information System data) and three dimensional displays. It is important when using quantitative data in

reports to present them in accessible and attractive ways – using pie-charts, coloured charts and diagrams is always preferable to a page full of figures.

Given the emphasis on bottom-up programmes and communication with a variety of stakeholders, written reports need to be supplemented by other media. For example, press releases, photographs, video footage, web-based communications and bulletin boards can all be used to improve communication and the accessibility of the evaluation data.

4.6 ACQUIRING AND USING DATA IN EVALUATION

All data is ‘produced’

Evaluators depend on something called 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’ - and data that are 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 and other data bases kept by DG REGIO.
- Annual reports of development authorities or federations of enterprises.
- Administrative records of public employment agencies, taxation returns, qualifications and training data .

Primary and secondary 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.
- 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 does 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 is 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 – that which is 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 is collected directly by the evaluation team its strengths, limits scope and relevance need to be thought through in terms of the kinds of future analyses that will be made and the kinds of arguments that the data will be expected to support.

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

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 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.

Accessing data can be difficult – it needs to be planned

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.,

There are ways to overcome problems of data access

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 as being included in the evaluation agenda and therefore as seeing 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, 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 and planning as an integral part of gathering data and choosing methods.

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 needs 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 strength 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. Categorical 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.

Quantitative data can be of different strengths

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 categorical, ordinal and interval data. Such measures cannot be considered 'objective' or as strong as it is sometimes assumed quantitative data always is. 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 in most evaluations in this area.

As the above distinctions suggest quantitative/qualitative data can be best understood as a continuum from the most quantitative to the most qualitative. In many ways categorical 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

A continuum from the quantitative to the qualitative

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. People's 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 is 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.

4.7 CREATING INDICATORS AND INDICATOR SYSTEMS

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.

Indicators measure objectives, effects, quality and context

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.
- It is measured on an independent basis. It is preferable that information is collected by agencies not directly responsible for the intervention or legislation.
- The measurement is based on reliable data.

Indicators should be closely linked to policy goals, measured regularly and independently and reliable

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.

Indicators are informed by different sources

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 quantifies, it should be easily communicated and understood.

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 indicators²:

- In relation to variables: Complete, partial, complex
- In relation to the processing of information: Elementary, derived and compound indicators
- In relation to the comparability of information: Specific, generic and key 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

Several typologies of Indicators exist. Resource, Output, Result, Impact and Contextual Indicators are most useful for socio-economic programmes

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

² These typologies are consistent with those in the document "Common Guidelines for Monitoring and Evaluation" (European Commission, 1995, Luxembourg: OPOCE).

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.

Resource indicators capture the inputs to the programme.

Output indicators represent the product of the programme's 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.

Output indicators measure what was produced.

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 easily known to the operators, so they are generally quantified exhaustively during monitoring.

Results 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.

Result indicators measure the advantages to beneficiaries

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.

Impact indicators measure the consequences of intervention.

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 a 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.

Impacts are obtained through market and non-market effects. They tend to be apparent only after the end of the 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?)

The use of indicators in evaluation

Indicators serve a number of useful roles in evaluation. Their use is common with respect to programme evaluation, particularly where objectives are expressed in clear operational terms. The use of indicators normally forms part of an evaluation. 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 indicators 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.
- Indicators can be used to identify what would have happened in the absence of the initiative, policy or legislation (the counterfactual).

Indicators help to inform evaluative judgements in a range of ways

The system of indicators and the programme cycle

Indicators are used at the beginning of the programme cycle to help in defining territories eligible for assistance, in analysing the regional context, in diagnosing economic and social problems to be addressed, and in

Indicators can highlight disadvantage

assessing 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.

and problems to be addressed

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. Box 4.2 provides an example.

Indicators should be quantified and relate to clear objectives

Box 4.2 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 and at the mid-term stage. 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.

Indicators are a way of summarising progress on programmes

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 help to identify factors which people can relate to

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 the European Structural Funds 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 (less than six) which break down the global objective into its main strategic dimensions;
- the measures level (from one to several dozen), corresponding to the basic unit of programme management. Each measure has its own specific management apparatus. (Actions, which correspond to the smallest homogeneous component in the programme – since each

action groups together similar projects – are considered on the same level as measures);

- the project level (often a few 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 be operate satisfactorily.

As far as the users are concerned, explicit support from the highest level of the authority managing the programme has to be assured. It is then advisable to create a group of future users of the system, and to give it the job of defining the indicators. The composition of this group will probably not be very different from that of the group that will have to conduct the evaluation.

Involvement of suppliers and users can improve the choice of 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 Box 4.3), 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.

Some indicators will attract public attention more than others

Box 4.3 – 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.

The suppliers of information will know the 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.

Indicators should be piloted initially before being implemented.

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.

Statistical data should be backed up by expert opinion

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 once³. When there are too many indicators the decision-makers are swamped with an excess of information.

Focus on a small number of indicators (up to 10) helps in decision making

The heterogeneity of programmes

The experience of the Structural Funds 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.

Often only the most relevant indicators are measured.

In practice, it is impossible to produce and regularly use such a large amount of information. For example, the initial proposal for a Structural Fund Programme in Burgundy (France) consisted of over 200 indicators. In the end, only about fifty of them were quantified.

The number can be limited by choosing generic indicators and

³ Innes de Neufville (1994)

grouping them

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 4.4). This does not mean, however, that additional indicators may not be required to meet the operators' information needs.

One approach for limiting the size of the systems of indicators, without losing relevant information, is to identify generic indicators, or to group together indicators by category of beneficiary.

Box 4.4 – 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 manipulability 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.

Colin Stutt Consulting, (1997)

Northern Ireland Single Programme 1994-99; Mid Term Review External Evaluation

Suggestions for limiting the size of systems of indicators

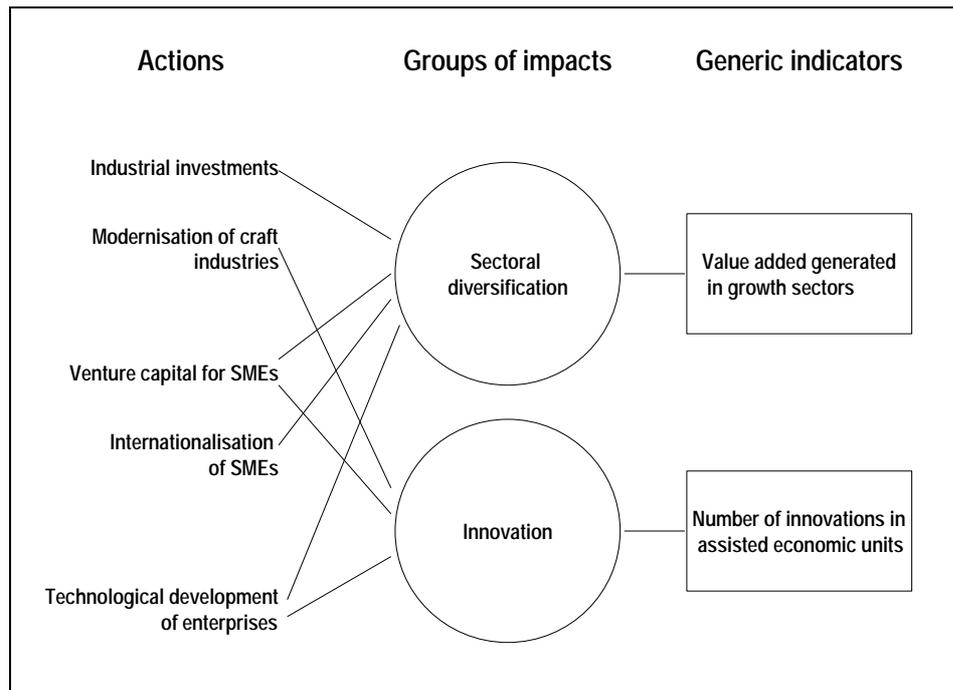
Suggestions for limiting the size of systems of indicators are typically based on the use of generic indicators or on the grouping of indicators by category of beneficiary. A lighter system limits the collection and circulation of information to the most essential elements at the programme level. On 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.

Finding generic impact indicators

Impact indicators are indispensable in evaluation but they are difficult to design and quantify. Apart from the number of jobs created, it is particularly rare to find generic impact indicators in programming

documents (some examples are given in Box 4.5).

Box 4.5 Examples of generic impact indicators



MEANS developed a method for creating generic impact indicators, which involved the following four steps:

- The evaluators gather the main documents relating to the programme (programming document, reviews, progress reports, brochures) and identifies all the sentences describing the objectives, the performance and the expected or real impacts.
- From the quotations taken from the documents, the evaluator selects those concerning impacts.
- It organises one or more seminars with the programme managers and operators. During these seminars, the participants complete the list of impacts and group them into families. Two techniques can facilitate this work: impact mapping and Metaplan.
- The participants of the seminar identify together the signification of each family of impacts, to give a name to each family and to choose corresponding indicators. They check that these indicators are generic, that is to say, that they can be applied to numerous actions within the programme.

Assessing the quality of a system of indicators

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 to recommending ways of enhancing it. 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. The availability of data to allow quantification is the primary factor to be considered. Monitoring indicators should be quantified at each monitoring meeting, that is to say, every six to twelve months. Evaluation indicators are quantified less frequently, typically annually or every three to six years.

The usefulness and quality of an indicator depends on: the availability of data; sensitivity to the intervention; reliability & credibility; comparability; normativity; meaning; and validity.

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 assisted businesses is an indicator that is not sensitive enough. 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.

Credibility

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.

Comparability

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.

Normativity

A good indicator must be understood by everyone who has to use it. In the

Validity

minds of both decision-makers and the public, the meaning of the indicator must be the same as for the programme managers. It must accurately reflect the concept to be measured. This is sometimes referred to as validity.

These criteria were applied to the assessment of the Portuguese programme document indicators for the "transport" sector for the period 1994-1999.

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. In particular, result and impact indicators should be the most numerous.
- 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.

Indicator systems should provide: coverage; balance; selectivity; and relevance.

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. For example the indicators used for monitoring the Structural Funds should be consistent with those used for monitoring the European Employment Strategy and the Social Inclusion Strategy.

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. There have been numerous opportunities for this type of comparison provided by the INTERREG Programme (see Box 4.6).

Box 4.6 Cooperation to generate transnational comparisons

Pays-de-la-Loire (France) and Emilie-Romagne (Italy) regions held bilateral meetings in 1997-1998 to discuss their evaluation work in the domain of rural tourism. During two two-day seminars, the regions analysed their respective actions and jointly defined a logic diagram of outputs, results and impacts. Common indicators were proposed, with a view to making inter-regional comparisons.

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 4.7). Although mainly concerned with contextual indicators rather than evaluation the Urban Audit provides an example of efforts to achieve comparability.

Box 4.7 – 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 Office* developed a series of standard indicators. Some of these indicators concern the context: rate of employment and unemployment; productivity; manpower; 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.

Proposals for key publicly accessible indicators

Box 4.8 proposes a series of indicators that might be used as key publicly accessible indicators. This list is neither closed, nor directive, nor stable and is limited to programme indicators. The indicators have been given scores, from *** to *, in decreasing order of interest. The indicators concern: Resources; Outputs; Results; and Impacts.

Box 4.8 Proposals for key indicators: Resources

Interest	Indicator
	Human resources
**	Temporary employment in the firms undertaking works during implementation (jobs x years)
*	Number of operators (public and private organisations responsible for providing assistance to beneficiaries)
*	Number of advisors (FTEs) mobilised to provide advice to beneficiaries
	Financial resources
***	Rate of budget absorption (% of allocated funds)
**	% projects (in financial terms) especially benefiting women
**	% projects (in financial terms) in rapidly growing markets / sectors
*	% of budget devoted to environmental mitigation measures
*	% projects (in financial terms) concerning the most disadvantaged areas

Outputs

Interest	Indicator
	Progress of works
***	Rate of completion (% of objective)
**	Compliance with project duration
	Capacity of finished works
**	Number of potential connections (business / households) to networks of basic services (broken down by services)
	Activity of the operators in terms of attracting and selecting participants
**	Selection rate (% of projects accepted as a proportion of eligible

	projects)
**	Coverage rate (Penetration): % of the target population who have been (should be) participants in the programme
*	% of beneficiaries belonging to priority groups (e.g. long-term unemployed, early school leavers)
*	% of beneficiaries situated in the most disadvantaged areas
**	% of beneficiaries involved in rapidly growing markets / sector
***	% of women in beneficiaries
***	% of SMEs in beneficiaries
	Services funded by the programme
***	Number of individual beneficiaries having received services, advice, training
***	Number of economic units (enterprise, farm, ship owner, fish farm, tourism professional) having received services, advice, training
**	Number of hours of training / advice provided to beneficiaries

Results

Interest	Indicator
	Satisfaction of beneficiaries
*	Satisfaction rate (% of beneficiaries that are satisfied or highly satisfied)
	Benefits gained by beneficiaries
**	Average speed between principal economic centres
	Investments facilitated for beneficiaries
**	Leverage effect (private sector spending occurring as a counterpart of the financial support received)

Impacts

Interest	Indicator
	Sustainable success
**	Rate of placement (e.g.: % of individual beneficiaries who are at work after 12 months, incl. % in a stable long-term job)
*	Rate of survival (e.g.: % of assisted economic units that are still active after 12 / 36 months)
	Impact perceived by beneficiaries
***	Value added generated (e.g.: after 12 months in terms of euros / year / employee,)
***	Employment created or safeguarded (e.g.: after 12 months Full Time Equivalent)
	Impact globally perceived in the area
**	Residential attractiveness (e.g.: % of inhabitants wishing to remain in the area)
	Indirect impact
*	Regional knock-on effects (e.g.: % of regional firms within the suppliers of assisted firms after 12 months)

4.8 USING INDICATORS TO IMPROVE MANAGEMENT

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 administration⁴.

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.

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 4.9).

*Comparison:
planned with
actual*

Box 4.9 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 disqualified from the programme at each step of its 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.

Whenever possible, it is useful to compare programme indicators (for example, number of unemployed trained) with the appropriate context

⁴ For example, the US "Government Performance and Results Act (GPRA)" of 1993 made the use of new indicators compulsory through the entire government administration.

indicators (for example, number of unemployed 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 (see Box 4.10).

Box 4.10 CSF (Objective 1) 1989-93 – Comparison of programme and context indicators – Spain

	Recorded data	Data as a percentage
Number of digital telephone lines	239 800	6 % of the digital lines installed in Spain during the period
Funded universities	20	91 % of the Universities in the Objective 1 regions
Creation and improvement of irrigated land (in ha)	34 236	1.5 % of the surfaces irrigated in the Objective 1 regions
Number of modernised farms	107 000	4.8 % of the farms in Spain (1992)
Number of young farmers subsidised	11 000	0.5 % of the farms in Spain (1992)
Number of persons trained	2 693 000	17.7 % of the working population in Spain

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.

Comparison: programme and context

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 4.11).

Box 4.11 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). 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

Comparison: one intervention against another

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 given in Box 4.12. 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".

Box 4.12 Example of comparison of performance accepted by the operators

The comparison of performance between training organisations often stumbles over the fact that the different publics trained do not have the same difficulties. A few years ago an evaluation team had to compare 133 training organisations (school districts) in different states in the U.S. The team decided to compare each organisation to a "benchmark group" of 14 other organisations with similarities as regards the characteristics of the public enrolled (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 performance comparisons were accepted without question.
Henry et al, (1992)

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 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" potential direct beneficiaries by favouring those whose employability is higher. This effect is undesirable because it helps to focus assistance on those who are relatively less in need.

Indicators can cause perverse behaviour

An example of how indicators caused a reduction in differences by a convergence towards the average is given in Box 4.13.

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.

Box 4.13 Adverse affects of an 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 probable 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.

Over time, adverse effects may be inevitable, but are not an argument for not measuring performance

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 Box 4.14).
- Operators with the best performance are offered additional funds.

Indicators can be used to improve performance when accompanied by incentives

Box 4.14 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, the services of the Regional Council applied a project selection method used for their own training activities. This method is based on performance management and functions in the following way.

The training organisations filing their first application for funds are subjected to less

severe selection procedures, on the basis of the quality of their projects. This selection mechanism is exceptional since "primary applications" are rare.

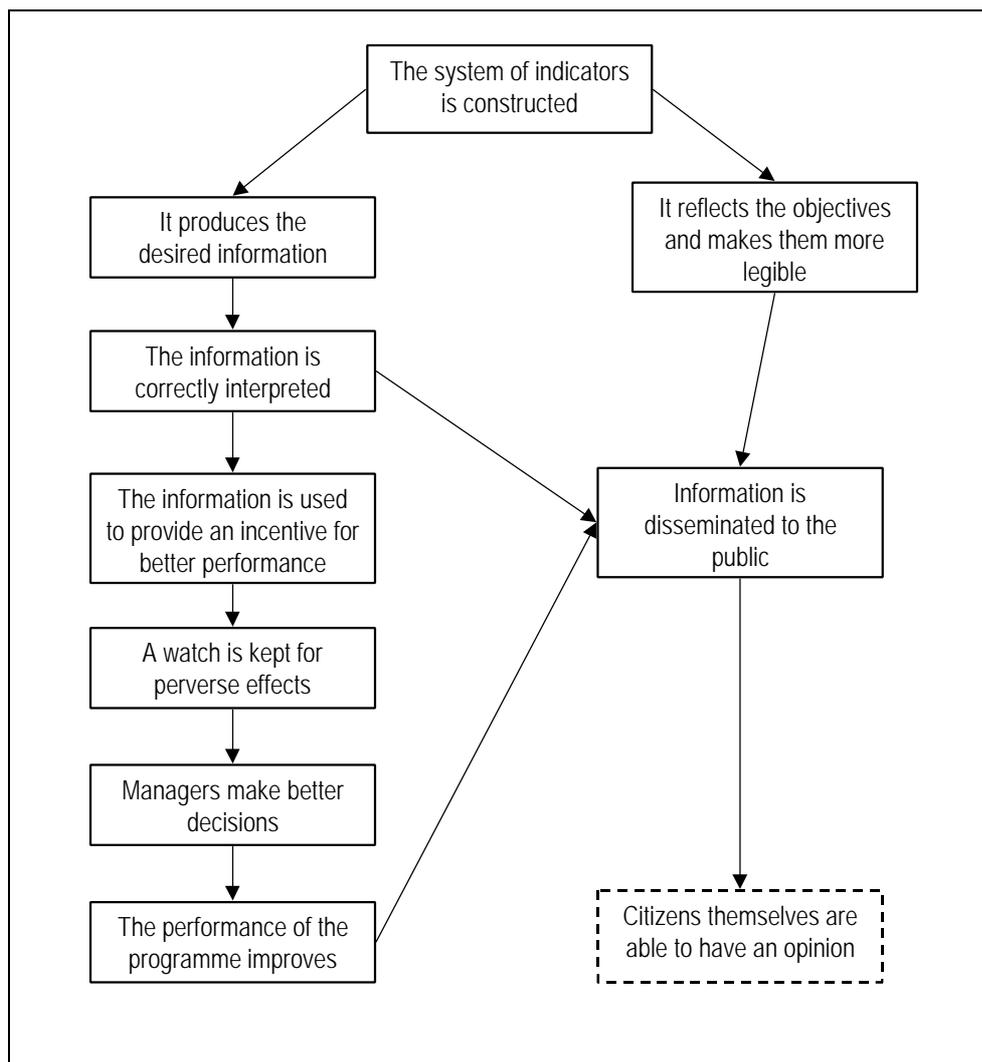
Organisations already receiving assistance and which apply for a new subsidy (secondary applications) enter into the mechanism of normal project selection. In this normal procedure, funds are granted first to organisations with the best placement rate. The monitoring system obliges all organisations that offer subsidised training to provide the Regional Council with an evaluation of the placement rate of its trainees after six months.

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.

The usefulness of a system of indicators

Box 4.15 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.

Box 4.15 A system of Indicators



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.

4.9 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 is 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 on 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 judgments. 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 is never pure or naturally occurring, it needs to be produced. Because of this evaluators need to know where their data comes and what decisions have made in the course of its production. At the end of the day the strength of the arguments and conclusions that can be drawn depend on the strength and characteristics of the data that is being used.
9. One important aspect in quality in evaluating data follows from the way data has 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 that 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 is 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 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 monitoring 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 a 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 as well as their management. This would appear to be the case in some countries where the performance reserve has been positively incorporated into programmes within the European Structural Funds.

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 final 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.

Mid-term 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.

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.

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.

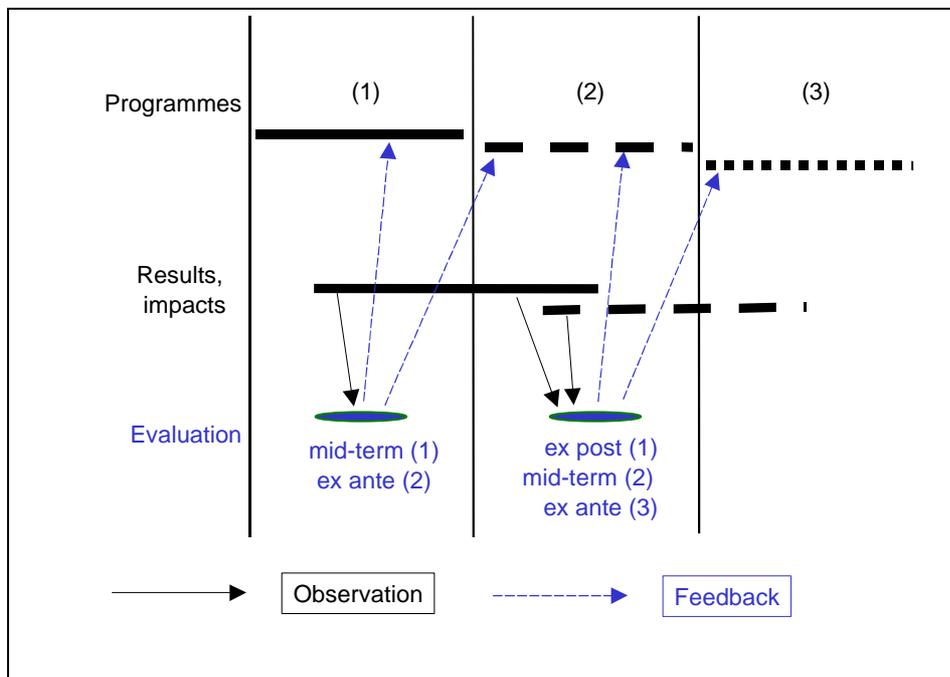
Ideally, the results of this evaluation should be available when the next programme is planned, that is, at least a year before the end of the programme. However, for the impacts to have been produced, ex post evaluation would have to be performed two to three years after the end of the programming period. 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 three evaluation phases in successive cycles creates overlaps that have to be organised as efficiently as possible to avoid any duplication of work. The basic principle is that of combining evaluation work, during a programme, with the use of conclusions of evaluation performed on the preceding programme. The relative continuity of actions programmed from one period to the next makes it possible to use conclusions from the recent past to judge the relevance of the new measures proposed. The following diagram shows that interactions are possible between evaluation work performed at the different phases of several successive programmes.

Box A.1 – Articulation of evaluation programming cycles



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

- 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.
- 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 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 pluri-annual evaluation plan. The idea is to identify the different possible evaluations and to establish their date and content in relation to the political schedules and deadlines for decision-making of the evaluation "customers" at various levels.

Annex B: Changes in Structural Fund regulations

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 current generation) included specific articles on evaluation, particularly Article 6 of the framework regulation and Article 26 of the co-ordination regulation.

Regulations applicable between 1988-93

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 pluri-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.

Regulations applicable between 1994-99

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 particularly successful, as our

survey below shows. 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.

Regulations of the third generation of programmes (2000-2006)

With the third generation of Structural Funds the assessment of their effectiveness will be reinforced. Community actions will henceforth be the object of ex ante evaluation by the Member States themselves, of intermediate 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 is 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, may be launched with a view to identifying experiences that are transferable from one programme to another.

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

Annex C: References

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: état 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 Brunswick & London: 2001

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

Evaluating Socio Economic Development, The GUIDE: Annexes

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