



European Commission
Enterprise
Directorate General

Österreichisches Institut
für Gewerbe- und Handelsforschung
Austrian Institute for Small Business Research



Evaluating Actions and Measures Promoting Female Entrepreneurship

A Guide



The responsibility for the content of this report lies with the
Austrian Institute for Small Business Research (IfGH)
A-1040 Vienna, Gußhausstrasse 8
tel. +43-1 505 97 61, fax +43-1 503 46 60
Internet: <http://www.ifgh.ac.at>, e-mail: ifgh@ifgh.ac.at



EUROPEAN COMMISSION

**Evaluating Actions
and Measures Promoting Female
Entrepreneurship**

A Guide

Austrian Institute for Small Business Research (IfGH)

Authors:

Sonja Sheikh (IfGH)

Nadia Steiber (IfGH)

Vienna, December 2002

PREFACE

This guide is an outcome of the study *'Examination and Evaluation of Good Practices in the Promotion of Female Entrepreneurship'* commissioned by the European Commission, Enterprise Directorate-General. The study was carried out by the Austrian Institute for Small Business Research (Österreichisches Institut für Gewerbe- und Handelsforschung, IfGH) in Vienna from January to December 2002 in co-operation with members of the European Network for SME Research (ENSR) and other partner institutions from all Member States of the European Union and the EFTA/EEA countries.

In addition to this guide for *'Evaluating Actions and Measures Promoting Female Entrepreneurship'* prepared by Sonja Sheikh and Nadia Steiber, a publication on *'Good Practices in the Promotion of Female Entrepreneurship – Examples from Europe and other OECD Countries'* and a *'Female Entrepreneurship Database'* (MS ACCESS) have been produced in the framework of the above mentioned project. The publication on good practices in the promotion of female entrepreneurship provides an overview of specific actions and support measures adopted by national administrations in the Member States of the European Union and the EFTA/EEA countries in order to promote female entrepreneurship, particularly in the fields of start-up, funding, training, mentoring, information, advice and consultancy, and networking and presents selected cases of good practice from Europe as well as from other OECD countries. The female entrepreneurship database provides detailed information on actions and measures promoting female entrepreneurship in the Member States of the European Union and the EFTA/EEA countries, particularly on the organisation of their contact points, their contact details, their content, their objectives, their source of funding, their operational elements, etc.

The publication on *'Good Practices in the Promotion of Female Entrepreneurship – Examples from Europe and other OECD Countries'*, the *'Female Entrepreneurship Database'* and the present evaluation guide may be obtained from the European Commission, Enterprise Directorate-General, at the following Internet address: <http://europa.eu.int/comm/enterprise/entrepreneurship/craft/craft-women/women-dgentr-activities.htm>. Printed versions of this guide and of the publication can be requested by e-mail to the following address: Entr-Craft-Small-Business@cec.eu.int or by fax: +32 / 2 / 299 81 10.



CONTENT

1	INTRODUCTION.....	1
2	USEFUL DEFINITIONS AND REASONS TO EVALUATE	3
2.1	What is an Evaluation?.....	3
2.2	Reasons to Evaluate Programmes Promoting Female Entrepreneurship.....	5
3	PREPARING AND MANAGING AN EVALUATION.....	7
3.1	Preparing an Evaluation.....	7
3.1.1	Defining the Purpose and Scope of Evaluation.....	7
3.1.2	Determining the Right Moment for Evaluation.....	8
3.1.3	Knowing the Main Evaluation Criteria	10
3.1.4	Formulating Evaluation Questions	12
3.2	Managing an Evaluation.....	13
3.2.1	Deciding for an Internal or External Evaluator	13
3.2.2	Drawing up the Terms of Reference	14
3.2.3	Steering an Evaluation	17
4	CONDUCTING AN EVALUATION.....	19
4.1	Evaluating the Context and Intervention Logic.....	20
4.1.1	Analysing the Context of a Programme	20
4.1.2	Analysing the Intervention Logic of a Programme	21
4.1.3	Judging the Relevance of a Programme.....	24
4.1.4	Judging the Coherence of a Programme.....	25
4.2	Analysing the Implementation of a Programme.....	26
4.2.1	Monitoring the Implementation of a Programme.....	26
4.2.2	Evaluating the Implementation of a Programme.....	29
4.3	Evaluating the Effects of a Programme	31
4.3.1	Choosing a Suitable Evaluation Design.....	31
4.3.2	Selecting Appropriate Indicators.....	33
4.3.3	Applying Suitable Techniques for Data Collection	35
4.3.4	Estimating Net Effects of a Programme.....	39
4.3.5	Judging the Effectiveness of a Programme	41
4.3.6	Judging the Efficiency of a Programme	42
5	ASSESSING THE QUALITY OF AN EVALUATION	45
6	DISSEMINATING AND USING EVALUATION RESULTS	51





1 INTRODUCTION

Promoting equality for women in business is one of the central aims of most programmes targeted at the promotion of female entrepreneurship. It means redressing existing imbalances between the position of male and female entrepreneurs and advancing the status of women entrepreneurs to a level where they can participate in and benefit from developments on an equal footing with men. Effective promotion of female entrepreneurs will have to take account of the specific needs of women and contribute to tackling the problem of different baselines for women and men aiming to start up or successfully run a business by providing tailor-made support. Virtually all Member States of the European Union and many other countries inside and outside Europe have recognised the importance of promoting entrepreneurship in recent years and have implemented specific actions or measures to particularly support female entrepreneurship. However, there has never been a process of exchanging information or good practice between the Member States in this field. This is why the European Commission, Enterprise Directorate-General launched in 2001 a Best project on “Promoting Entrepreneurship amongst Women”. The Best project has been implemented in the framework of the “Multiannual Programme for Enterprise and Entrepreneurship and in particular for Small and Medium-sized Enterprises (SMEs) (2001-2005)”. It is aimed at collecting information on specific actions and measures promoting female entrepreneurship and at identifying good practices with the view to a possible future benchmarking exercise. Evaluating programmes targeted at the promotion of female entrepreneurship can constitute an important first step towards this direction.

Evaluation is a wide-ranging concept and at a general level virtually anything can be evaluated. In practice, however, the term evaluation is applied specifically to public sector interventions. Whereas private sector organisations have little need to justify or account for their actions, since they are judged by the market, public sector organisations are obliged to evaluate, in order to check the *raison d'être* of a public intervention, to confirm reproducible success stories and failures not to be repeated, and to report back to citizens. Thus, evaluation is an essential part of modern public sector management practice and is directly relevant to the effective use of regional, national and Community resources. If evaluation is well conducted and if the results of evaluation are used by decision-makers, it can contribute to improving public support, to increasing transparency of public spending and to enhancing accountability.

Evaluation of public interventions may be performed at three levels of decision-making: the project, programme or policy level. The term *project* refers to a single, non-divisible intervention with a fixed time schedule and dedicated budget. A *programme* signifies a co-ordinated set of different types of actions or measures (e.g. training courses, personalised advice, start-up support, etc.) directed towards the achievement of an objective or a set of objectives in a given period of time and limited in terms of budget. Finally, the notion of *policy* refers to a set of programmes and/or measures which have the same general objective or goal, but not necessarily the same specific objectives, beneficiaries and modes of management. Unlike projects and programmes, a policy is usually not limited in terms of time schedule or budget.

The present guide focuses on *programme evaluation* with a particular emphasis on specific actions and measures targeted at the promotion of female entrepreneurship in different fields, such as start-ups, training, advice and consultancy, mentoring, information, funding, and networking. However, most of the issues related to the evaluation of programmes promoting female entrepreneurship in these fields as well as in general, may be equally relevant for evaluations in other domains (e.g. small and medium-sized enterprises, research and technological development, environment, employment, etc.). Thus, most of the concepts dealt with in this guide may be of interest to those concerned with the evaluation of other support programmes as well.



Obviously, the process of planning and conducting an evaluation is different for every programme. There is no single right way of doing evaluation. Each programme serves a different mix of clients, uses different service delivery approaches, defines different outcomes, is at a different phase of development, and faces different contextual issues. Therefore, the process to be applied for evaluating a specific programme targeted at the promotion of female entrepreneurship will depend to a large extent on local conditions and circumstances. However, a set of methodological issues exists which permit the assessment of the performance of a selected individual programme over time and which are common to all programmes targeted at the promotion of female entrepreneurship.

This guide addresses programme managers, rather than evaluation specialists. It shall introduce programme managers or other persons responsible for the implementation of programmes or measures targeted at the promotion of female entrepreneurship to the main aspects related to steering an evaluation (i.e. preparing and managing an evaluation, defining the fields of investigation, outlining the methods to be used for evaluation, assessing the quality of the work, and supervising the use of evaluation results) and shall provide a broad overview of the most relevant technical issues in this respect. At the same time the guide shows what can and what cannot be expected from an evaluation at a particular time and in a specific context. Thus, this guide is not intended to serve as an exhaustive instructional manual for conducting evaluation, but provides a framework for thinking about evaluation as a relevant and useful programme tool and for commissioning and supervising evaluations of programmes targeted at the promotion of female entrepreneurship.

The next section of this guide introduces some useful definitions related to the term evaluation - particularly in contrast to audit and monitoring - and elaborates three good reasons to evaluate programmes targeted at the promotion of female entrepreneurship. In section 3 some basic issues referring to the preparation and management of evaluations are described, ranging from the definition of the scope and purpose of an evaluation, the selection of evaluation criteria and the formulation of evaluation questions to drawing up the terms of reference and steering an evaluation. Section 4 deals with main concepts related to the conduction of an evaluation, focussing on the evaluation of the context and intervention logic of a programme, the monitoring and evaluation of programme implementation and the evaluation of the longer term impacts of a programme or measure promoting female entrepreneurship. In section 5 a grid for assessing the quality of an evaluation is provided and, finally, section 6 contains some basic considerations on the dissemination and use of evaluation results.



2 USEFUL DEFINITIONS AND REASONS TO EVALUATE

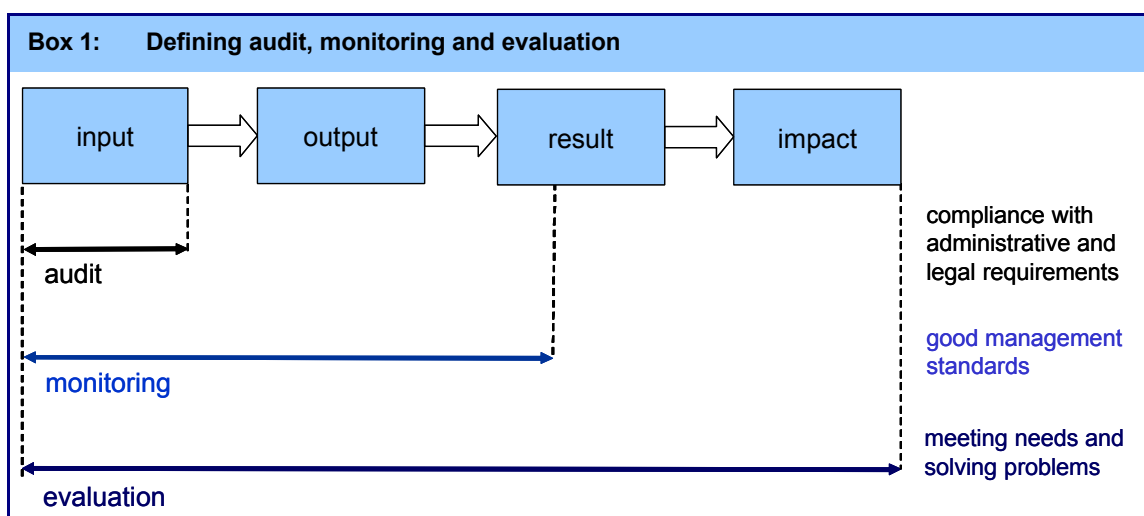
Before going into detail about how to prepare and manage an evaluation and how to perform it, it seems to be necessary to have a clear understanding of what an evaluation is and why evaluating programmes or measures targeted at the promotion of female entrepreneurship might be useful. In order to ensure that the expectations placed on evaluation do not exceed its real contribution and to avoid assigning roles to evaluation which it is not designed to play, this chapter provides some basic definitions for the three similar concepts of audit, monitoring and evaluation and elaborates three good reasons for evaluating actions, measures or programmes aimed at the promotion of female entrepreneurship.

2.1 What is an Evaluation?

There is no general agreement on what evaluation is. This might be explained by the variety of disciplines (economic, policy and administrative studies, statistics, sociology, psychology, etc.), institutions and practitioners active in evaluation and the wide range of issues, needs and clients that are served by evaluation. A number of different definitions of the term evaluation have been put forward by various authors. Summarising these definitions, an evaluation can be considered as:

The judgement of a public intervention according to its results, its impacts and the needs it intends to satisfy.

The concept of evaluation must be clearly distinguished from the related disciplines audit and monitoring. These three exercises are similar in so far as they are judgements of public actions and can mutually enrich each other, but they require different professional qualifications, time frames and modes of organisation. An initial clarification concerns the point of view from which a public intervention is judged. *Audit* verifies the legality and regularity of the implementation of resources. *Monitoring* verifies the sound management of the intervention and produces a regular analysis of the progress of outputs. *Evaluation* judges the implementation of the intervention on the basis of the outputs, results and impacts it has produced in society and ultimately whether it has met the needs and solved the problems it was supposed to address. A second distinction between these three concepts concerns the judgement criteria. *Audit* judges in terms of criteria that are known and clarified in advance (budgets, regulations, professional standards), *monitoring* judges in terms of the operational objectives to be met, by contrast, *evaluation* often has to choose its judgement criteria when starting the exercise (see section 3.1.3 for an overview of the main judgement criteria applied to evaluation). These basic differences distinguishing between audit, monitoring and evaluation are presented in Box 1 (see section 4.1.2 for a definition of the terms input, output, result and impact).



Summarising the above, the following definitions may be derived for the terms audit, monitoring and evaluation:

Audit

An audit judges an intervention according to the *inputs* or resources which have been implemented. It checks whether these conform to existing fixed standards, such as legal provisions, budgets and professional rules. Audits are usually conducted by auditors who have in-depth knowledge of pre-established criteria (legal or accounting knowledge, for example).

Monitoring

Monitoring judges an intervention according to its *outputs*. It continuously tracks performance against what was planned by collecting and analysing data on the indicators established for monitoring purposes and provides continuous information on whether progress is being made towards achieving results through record keeping and regular reporting systems. It is normally the programme managers or those responsible for implementing the intervention who conduct monitoring.

Evaluation

Evaluation judges an intervention according to its *results* and *impacts*. It is a periodic, in-depth analysis of the performance of an intervention and relies on data generated through monitoring as well as on information obtained from other sources (e.g. studies, research, in-depth interviews, focus group discussions, surveys etc.). Evaluation is typically (but not always) conducted with the assistance of external evaluators.

Although the underlying guide primarily deals with *evaluation*, the issue of *monitoring* is briefly dealt with in section 4.2.1, as monitoring plays an important role in the analysis of programme implementation.



2.2 Reasons to Evaluate Programmes Promoting Female Entrepreneurship

An evaluation of programmes and measures targeted at the promotion of female entrepreneurship is usually conducted with the general aim of improving the intervention. It may also be undertaken with the intention of identifying the programme's effects on (potential) female entrepreneurs and on society as a whole, or to allow decision-makers to judge the programme's value. However, beyond this rather general reason there are three specific reasons to evaluate programmes aimed at promoting female entrepreneurship.

◆ Verifying that the programme replies to uncovered or insufficiently satisfied needs

One of the aims of evaluation is to verify the *raison d'être* of a public action. A public action can be justified only if it is intended to satisfy prevailing needs or solve socio-economic problems. Evaluation of programmes or measures targeted at the promotion of female entrepreneurship, for example, may serve to clarify the needs of or to verify the existence of problems faced by (potential) female entrepreneurs, especially, if it is conducted *ex ante*. In this context it is also essential to rule on the appropriateness of the programme as such. When evaluation is conducted *ex post*, it can serve to examine whether the needs or problems of (potential) female entrepreneurs still exist (see section 3.1.2 for a definition of the terms *ex ante*, intermediate and *ex post* evaluation).

◆ Improving the programme

By observing the implementation, as well as the results and impacts of a public intervention, the evaluation exercise provides an opportunity for feedback arising from the action underway. Evaluation, and particularly intermediate evaluation, of programmes promoting female entrepreneurship may, thus, also be undertaken for managerial reasons, i.e. concerned with assessing and improving programme implementation. Typically, those involved in managing a programme or measure promoting female entrepreneurship need to know what its strengths and weaknesses are, how it can be improved, which aspects of the action work adequately and which aspects do not, and what are the reactions of clients, staff and others to the intervention. Evaluation, and particularly *formative* evaluation, can contribute to the learning process of programme managers and operators by analysing the factors of success or failure of a programme and helping them to adjust the modalities of the intervention and its implementation (see section 3.1.1 for a definition of formative and summative evaluation).

◆ Enhancing accountability

Finally, evaluation may also be conducted for reasons of accountability, transparency or liability by reporting to political authorities and citizens on the results obtained and on the sound use of the resources allocated to an intervention. Evaluation, and to a great extent *summative* evaluation, helps to explain in simple terms where public money is spent, what effects it produces and how the spending is justified. Hence, it can strengthen the legitimacy of programmes and secure the sustainability of results. If accountability is at the forefront, the evaluation of a programme or measure promoting female entrepreneurship is likely to focus on the issue of effectiveness as reflected by empirical evidence and in the perception of the main stakeholders, as well as on possible side-effects.





3 PREPARING AND MANAGING AN EVALUATION

As evaluation studies are limited by available human and financial resources as well as by the time available to plan, conduct and disseminate evaluation findings, it is very important to ensure that the evaluation is well prepared and properly managed - irrespective of whether it is conducted internally or externally - so that it is credible and provides maximum use for the resources spent on it. In this chapter more detailed information on how to effectually prepare and manage an evaluation is provided.

3.1 Preparing an Evaluation

When an evaluation is not well prepared, there is the danger that it may be carried out inefficiently. With regard to the preparation of an evaluation it is particularly important to precisely define the aim and purpose of the evaluation, to choose evaluation criteria according to which the evaluation shall take place, to determine the right moment for the evaluation and to formulate evaluation questions that shall be answered by the evaluation.

3.1.1 Defining the Purpose and Scope of Evaluation

The scope

The scope of an evaluation specifies which programme, action, or measure or which part of it should be covered by the evaluation and what aspects of the evaluation are to be considered in which depth. The field of investigation should be delimited in institutional, temporal, sectoral and geographical respect. Subsequently, defining the scope of an evaluation amounts to asking:

What is going to be evaluated?

Thus, the scope of an evaluation is a question of the breadth of the exercise and its depth. Whereas the *breadth* of the evaluation strongly depends on the type of intervention the evaluation is intended to cover, the *depth* is essentially determined by the evaluation criteria applied (see section 3.1.3 for an overview of the main evaluation criteria). Some of the most important questions to ask when defining the scope of an evaluation are:

- ◆ *Shall the programme be evaluated in isolation, or shall links between the programme and other related interventions at European, national, or local level be examined?*
- ◆ *Shall the evaluation be limited to the programming cycle under consideration or shall it include to a certain extent preceding cycles?*
- ◆ *Shall the effects of the programme be evaluated with regard to society as a whole or to particular social groups (e.g. female entrepreneurs) only?*
- ◆ *Shall the evaluation be limited to the eligible area of the programme or shall observations be extended to certain neighbouring or other areas which encounter similar problems or needs?*

The specification of the scope of an evaluation shall always relate to the objectives of the programme or measure to be evaluated. However, both in individual measures as well as in more complex types of interventions, objectives might often be unclear or formulated inadequately. In this case, it has to be decided whether the evaluation should entail a redefinition of implicit objectives. Also, interventions may have many unforeseen impacts, both positive and negative ones – which may well be overlooked if evaluation only focuses on what has been formally agreed upon. The possibility to overcome such problems depends on the definition of the *terms of reference* (see section 3.2.3) and the experience of the evaluation team as well as its ability to define analytical frameworks for the investigation.



The purpose

To be effective, evaluation needs a clearly defined *purpose*. The question to be asked after having determined the scope of evaluation is:

For what purpose is the evaluation launched?

Shall the evaluation primarily provide decision advice, seek information or contribute to learning? How will the outcomes of the evaluation be used? Who are the intended users of the evaluation and what are the expected results? Answers to these questions will help the commissioner to identify the main purpose of the evaluation, which will in turn affect the specific evaluation questions to be addressed (see section 3.1.4 for guidance on the formulation of evaluation questions).

Usually, two types of evaluation can be distinguished, depending on its purpose: formative and summative evaluation. A **formative evaluation** examines ways of improving and enhancing the management, implementation and development of a programme. It is generally conducted for the benefit of those managing the intervention in order to help them improve their work. A formative evaluation is usually undertaken during the implementation phase of a programme (intermediate or on-going evaluation). A **summative evaluation** determines the essential outcomes of an intervention and judges it according to given criteria, such as effectiveness or efficiency, for example (see section 3.1.3 for a definition of the terms effectiveness and efficiency). It is usually conducted some years after finalisation of a programme (ex post evaluation) for the benefit of external actors, who are not directly involved in the management of the public intervention to assist in allocating resources or enhancing public accountability (see section 3.1.2 for a definition of intermediate and ex post evaluation).

3.1.2 Determining the Right Moment for Evaluation

As indicated above, evaluation can be conducted at different stages of a programme. Depending on the phase of the programming cycle at which the evaluation is performed, it differs in its form, scope and focus. Particularly, the following three types of evaluation have to be distinguished: *ex ante* evaluation, which is conducted during the designing phase of a programme, *intermediate* evaluation, which takes place during the implementation of a programme, and *ex post* evaluation which is ideally performed two to three years after the end of the programme. The three types of evaluation are described below.

Ex ante evaluation

Ex ante evaluation takes place during the planning and development phase of a programme, i.e. before the programme has been adopted. At this stage evaluation is mainly required to determine the needs of the potential beneficiaries of a programme as well as the nature and scope of the intervention to be established. Typical questions in the field of female entrepreneurship to be posed in the scope of an ex ante evaluation are:

- ◆ *Are the proposed strategy and objectives of the programme relevant with respect to the needs of (potential) female entrepreneurs?*
- ◆ *Are the objectives of the programme clear and coherent and is there coherence between the objectives and other public policies in the field of gender equality and/or entrepreneurship?*
- ◆ *Are the expected impacts realistic?*



Intermediate evaluation

Intermediate evaluation takes place when the programme is implemented – in the pilot phase of a programme. At this stage, evaluation is mainly needed to determine whether the programme reaches its target group(s), whether the intervention is implemented as planned, and whether the objectives of the programme are likely to be reached. The primary intent is to critically analyse first outputs and results as well as to assess the financial management of the programme and the monitoring system installed. Depending on the conclusions of the intermediate evaluation, adjustments may be made during the programming cycle. Thus, intermediate evaluation has a *formative* nature, i.e. it produces direct feedback that shall help improving the management and implementation of a programme. Typical questions to be asked in the scope of an intermediate evaluation include:

- ◆ *Do the objectives of the programme remain relevant and coherent or have there been changes in the context that require certain objectives to be changed?*
- ◆ *Are the central objectives of the programme in the process of being achieved?*
- ◆ *Are there unexpected effects to be observed?*

Ex post evaluation

Finally, ex post evaluation judges the entire programme and particularly its impacts. It is ideally conducted two to three years after the finalisation of a programme and should, however, be available when the next programme is planned. Ex post evaluation aims at accounting for the use of resources and at reporting on the effectiveness and efficiency of interventions as well as on the extent to which expected effects are achieved (see section 3.1.3 for a definition of the terms effectiveness and efficiency). It has a *summative* nature, i.e. it aims at collecting and presenting information needed for judging the programme's effects and its value. Typical questions to be posed in the scope of ex post evaluations are:

- ◆ *Have the formulated objectives been achieved?*
- ◆ *How effective is the programme?*
- ◆ *How efficiently is the programme implemented?*
- ◆ *Could more beneficial effects be produced by adopting other implementation methods?*
- ◆ *Are the outcomes and impacts achieved sustainable?*

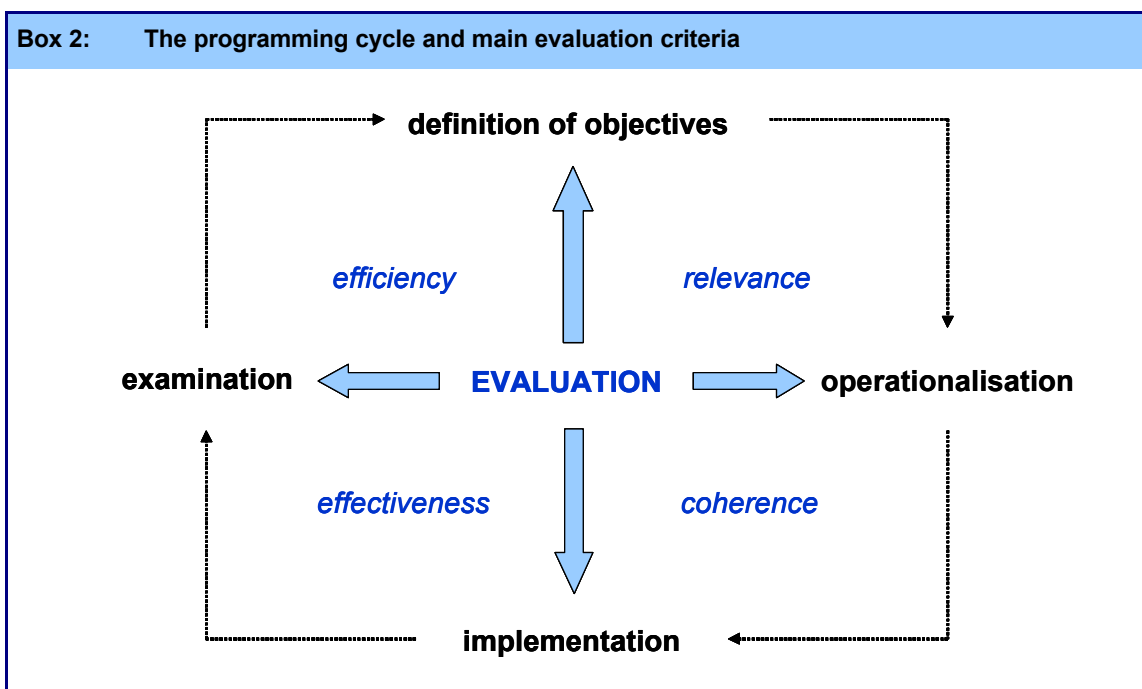
The underlying guide deals primarily with issues related to the conduction of intermediate or ex post evaluation of programmes targeted at the promotion of female entrepreneurship. However, some of the concepts presented are also relevant to ex ante evaluation.



3.1.3 Knowing the Main Evaluation Criteria

Evaluation is usually conducted according to a given set of evaluation criteria, whereby certain evaluation criteria predominate in evaluation theory: the *relevance* of a programme in relation to the needs of its potential beneficiaries, the *coherence* of a programme with other policy interventions in the respective field, the *effectiveness* of a programme in achieving its objectives, and the *efficiency* of an intervention, that is, its capacity to achieve the objectives at the lowest cost. An overview of these criteria is given in Box 2, schematically illustrating the different phases of a programme, starting with the definition of its objectives and running through its operationalisation, the implementation of its activities and its examination and assessment on the basis of the results and impacts it has produced.

As shown in Box 2, the relevance and coherence of a programme can and shall be judged before its implementation, whereas the effectiveness and efficiency can only be thoroughly analysed when the intervention has been running for a certain period of time.



A more detailed definition of the main evaluation criteria is given below:

Relevance

Relevance refers to the appropriateness of the explicit objectives of the programme in relation to the socio-economic problems and needs it is supposed to address. This evaluation criterion is most important in ex ante and in intermediate evaluation. A typical evaluation question in the field of female entrepreneurship that relates to the relevance of a programme is:

Is the programme justified with regard to the needs or problems of (potential) female entrepreneurs?



Coherence

Coherence refers to the degree of complementarity or synergy to be found within a programme and in relation to other programmes. The *internal coherence* refers to the correspondence between the different objectives of the programme. The *external coherence* refers to the adequacy between the programme to be evaluated and other interventions in the respective or related domains, e.g. actions supporting entrepreneurship or small and medium-sized enterprises. A typical evaluation question in the domain of female entrepreneurship relating to the coherence of a programme is:

Could better complementarity be achieved or more synergy be produced within the programme, compared to other public interventions in the field of equal opportunities and/or entrepreneurship?

Effectiveness

Effectiveness refers to the degree to which objectives are achieved and to which the problems targeted at by the programme have been resolved. This evaluation criterion is most important in ex post evaluation and to some extent in intermediate evaluation. A key evaluation question related to the effectiveness of a programme promoting female entrepreneurship may be:

In how far have the effects of the programme contributed to achieving its specific and its global objectives?

Efficiency

Efficiency refers to the degree to which economy is achieved while maintaining effectiveness. Efficiency is assessed by comparing the results obtained (preferably impacts produced) and the resources mobilised by a programme. This evaluation criterion is most important in ex post evaluation and to some extent in intermediate evaluation. A sample evaluation question related to the efficiency of a programme promoting female entrepreneurship is:

Could more beneficial effects be produced with the same resources or could the same results be obtained with fewer resources?

In addition to the above mentioned criteria, the criterion of *sustainability* may also be applied to evaluation, particularly to evaluation in the domain of female entrepreneurship.

Sustainability

Sustainability is concerned with what happens after a programme has been completed. Even if an intervention generates benefits which are in tune with the needs of its target population, it may be of little value unless these benefits are still being enjoyed at some stage in the future. For example, there is little use in training potential female entrepreneurs in skills which are likely to become obsolete after a few years. If a programme is to be of lasting value, it must generate sustainable benefits for its target population(s). A typical evaluation question that is related to the sustainability of a programme promoting female entrepreneurship is:

To what extent can the effects obtained be expected to last after the programme has been completed and no more public funding is provided?



3.1.4 Formulating Evaluation Questions

After having decided which evaluation criteria to apply (see section 3.1.3), these very general concepts have to be translated into concrete evaluation questions that specifically relate to the programme or measure under investigation. Probably the most fundamental aspect in the planning phase of an evaluation, and perhaps the biggest challenge, is to determine the questions to be answered by the evaluation. The specification of evaluation questions is particularly important in order to provide the evaluator with precise guidelines as to the exact information needs of the commissioners and stakeholders of the evaluation. These questions will obviously depend on the purpose and scope of the evaluation (see section 3.1.1) and on the phase of programme development, i.e. the timing of the evaluation (see section 3.1.2). However, critical evaluation questions to be addressed over the life of an intervention targeted at the promotion of female entrepreneurship may include, but are not limited to:

- ◆ *Is the programme justified with regard to the needs or problems of (potential) female entrepreneurs to be solved (relevance)?*
- ◆ *Does the programme reach the addressees it intends to serve?*
- ◆ *Could better complementarity be achieved or more synergy be produced within the programme, compared to related public interventions in the field of equal opportunities and/or entrepreneurship (coherence)?*
- ◆ *Which programme operations work well? Which ones do not work? Why or why not?*
- ◆ *Which strategies have been successful in encouraging participation and involvement? Which have been unsuccessful?*
- ◆ *What are the initial and long-term effects of the programme?*
- ◆ *In how far have the effects of the programme contributed to achieving its specific and its global objectives (effectiveness)?*
- ◆ *Does the programme have unintended consequences or unanticipated effects?*
- ◆ *Could more beneficial effects be produced with the same resources or could the same results be obtained with fewer resources (efficiency)?*
- ◆ *Which lessons might be learned from the programme that may be useful for further modifications, in case of a renewal or prolongation of the programme?*

All evaluations have multiple **stakeholders** who have an interest in the action being evaluated or in the results of the evaluation. Stakeholders may include sponsors, programme managers and administrators, participants or clients of the programme, regional or national authorities, collaborating agencies, and others with a direct or indirect interest in the programme. Ideally, the evaluation questions should be formulated based on the needs and interests of as many stakeholders as possible, in order not to miss any important question or issue that might be relevant for those who are not included in the process.



3.2 Managing an Evaluation

Once the decision to launch an evaluation has been made, effective management is crucial to the conduction of a successful evaluation. Several issues have proven to positively influence the effectiveness in this regard: the deliberate decision for an internal or external evaluation, the drawing up of detailed terms of reference, and the establishment of a system to monitor the evaluation exercise substantively. These issues are dealt with in more detail in the following sections.

3.2.1 Deciding for an Internal or External Evaluator

Once it has been decided which questions should be answered with the help of evaluation (see section 3.1.4), it will be easier to decide what type of evaluator is to be entrusted with the task of assessing the programme or measure under consideration. In general, three types of evaluators may be distinguished: external evaluators, internal evaluators and internal evaluators with an external consultant.

External evaluator

External evaluators are contracted from an outside agency or organisation to conduct the evaluation. As external evaluators maintain their position within their organisation, they generally have access to more resources than internal evaluators (i.e. computer equipment, support staff, library materials, etc.). In addition, they may have broader evaluation expertise than internal evaluators, particularly if they are specialised in programme evaluation and/or have conducted research in the respective domain of evaluation (i.e. on female entrepreneurship). External evaluators may also bring a different perspective to the evaluation as they are not directly affiliated with the programme under investigation. However, this lack of affiliation can be a drawback as well. External evaluators are not staff members; they may be detached from the daily operations of the programme and, thus, have limited knowledge of the programme objectives and its operational mechanisms, as well as limited access to programme activities. If an external evaluator shall be appointed, it should be noted that there are a number of different types of organisations which can perform an external evaluation. Two of the most frequently used types are:

◆ *Management consultancies*

Management consultancies vary from large, multinational firms which have considerable experience in carrying out a range of different evaluations, to smaller firms which possess a narrower, highly subject-specific expertise. Such firms are often perceived by stakeholders to embody a “businesslike” approach. Typically, these kinds of organisations can perform evaluations in a rather short time and tend to possess excellent presentation skills. However, they may also have disadvantages. Particularly, their prices may be comparatively high. The risk with management consultancies might be that they promise an evaluation but deliver an audit.

◆ *Academic institutions*

Academic experts are likely to offer a high degree of methodological expertise in evaluation. Some may also possess a high degree of subject specific knowledge. Stakeholders may tend to perceive academics as being relatively independent, which can be of advantage in circumstances where a management consultancy might be viewed with caution. An academic institution may represent better value-for-money compared to a management consultancy, but can often be less flexible. The risk with academic institutions might be that they promise an evaluation but deliver a scientific study.

If it is decided to appoint an external evaluator, she/he is usually chosen after a call for tenders (see section 3.2.2 for how to draw up the terms of reference).



Internal evaluator

A second option is to assign the responsibility for evaluation to a person or team already on staff or to hire an evaluator to join the programme. This internal evaluator might serve as both an evaluator and staff member with other responsibilities. As an internal evaluator works within the programme, she/he may be more familiar with the respective staff, will have access to organisational resources, and might have more opportunities for internal feedback with the programme's stakeholders. However, an internal evaluator may lack the outside perspective and technical skills of an external evaluator.

Internal evaluator with an external consultant

A final option combines the qualities of both types of evaluators. An internal staff person conducts the evaluation and an external consultant assists with the technical aspects of the evaluation and helps to gather specialised information. With this combination, the evaluation can provide an external viewpoint without losing the benefit of the internal evaluator's first-hand knowledge of the programme.

Irrespective of the choice for a specific type of evaluator, there are several criteria which an ideal evaluator of a programme or measure targeted at the promotion of female entrepreneurship should satisfy:

- ◆ specialist knowledge in the field of entrepreneurship, and particularly female entrepreneurship
- ◆ good knowledge in evaluation theory and methodology
- ◆ experience in the evaluation of entrepreneurship and SME development programmes
- ◆ expertise in gender issues and gender specific indicators
- ◆ independence and external legitimacy
- ◆ professional standing (i.e. ability to meet deadlines, communication skills, project management capacity, etc.)

3.2.2 Drawing up the Terms of Reference

The terms of reference are the document which serves as the basis of a contractual relationship between the commissioner of the evaluation and the team responsible for carrying out the work. Clearly defined terms of reference are vitally important when an evaluation is to be conducted by an external expert. They can, however, be of equal use when the evaluation will be performed internally. The terms of reference allow the commissioners of the evaluation to define their requirements and allow the evaluator to understand clearly what is expected of the work to be undertaken. They need to be sufficiently precise so that several evaluation teams can compete and the work may be performed in good conditions. The drawing up of the terms of reference therefore requires particular care. In Box 3 a standard layout of the terms of reference is provided. Adequate time should be invested in formulating and refining the terms of reference and key stakeholder of the programme or measure to be evaluated should be involved in the process as much as possible so that their views and interests can be reflected.



Box 3: Standard layout of the terms of reference

1. regulatory framework and motivation for the evaluation
2. scope of the evaluation
3. future use and users of the evaluation
4. evaluation questions and criteria
5. available knowledge and data
6. methodology for data collection and analysis
7. work schedule
8. suggested budget
9. required qualifications
10. structure of the proposal

1. Regulatory framework and motivation for the evaluation

The terms of reference must specify the legal and contractual requirements upon which the evaluation will be based, who initiated the evaluation project and, where relevant, who was involved in formulating the order for the evaluation. Also the motives for the decision to evaluate shall be explained, if they are not purely regulatory.

2. Scope of the evaluation

The terms of reference shall describe the scope and purpose of the evaluation (see section 3.1.1) and may also define its limits.

3. Future use and users of the evaluation

In order to enhance the future utility of the evaluation as much as possible, evaluators need to know how the findings of the evaluation will be put to use, who the primary intended users are and what results are expected from the evaluation.

4. Evaluation questions and criteria

The terms of reference shall pose a small number of clear and concise questions, to which the evaluation team shall provide answers in its conclusions (see section 3.1.4 for guidance on the formulation of evaluation questions). The core evaluation questions usually integrate the main or at least some of the main evaluation criteria (see section 3.1.1 for an overview of the main evaluation criteria).

5. Available knowledge and data

The terms of reference shall specify the form of existing knowledge (i.e. raw data, information structured in a computerised database, data already analysed, etc.) and may mention what the existing sources of other relevant information are and whether comparison references are already available. They shall also indicate who is responsible for gathering information (e.g. information is made available to the evaluation team by a particular actor or information is to be collected by the evaluation team itself).



6. Methodology for data collection and analysis

Without determining the details of the methods to be used for evaluation, the terms of reference must provide essential information for applicant teams to make reasonable methodological proposals. They should, for example, include information on:

- ◆ the required depth of the evaluation
- ◆ the priority evaluation questions
- ◆ the most important indicators from the commissioner's point of view

7. Work schedule

The terms of reference shall include factors, such as the duration of the evaluation, deadlines for respective reporting, and procedures to be followed to disseminate and use the evaluation results.

8. Suggested budget

If possible, the evaluation budget should be established by the commissioner considering the interest and difficulty of the questions asked. Generally, the budget for an evaluation shall amount to between 2 % and 5 % of the programme's total budget.

9. Required qualifications

The terms of reference shall specify the knowledge and experience required from the evaluation team (see section 3.2.1 for criteria an evaluator of a programme promoting female entrepreneurship should satisfy).

10. Structure of the proposal

To facilitate the selection of the evaluation team, the terms of reference may oblige the applicant team to organise their proposals in a certain way, for example:

1. background and essential elements of the evaluation context
2. composition and description of the evaluation team (including detailed information on previous experience)
3. evaluation method proposed
4. detailed work plan, including time schedule, organisation of work and distribution of tasks within the team.
5. price of the evaluation
6. technical, legal and financial references of the contracting organisations and individual references of the team members

The evaluation team shall be selected by taking into account its independence and profile and the methodological qualities as well as the quality-price ratio of its proposal. Possibly no one of the applicants can fully satisfy the requirements mentioned in the terms of reference. In this case, choosing an evaluator might involve compromising on one or the other point.



3.2.3 Steering an Evaluation

The steering of an evaluation may either be entrusted to a single official or a small group of officials from the administration most directly concerned with the programme or be exercised by a steering group created for this purpose and including all significant stakeholders who have a specific interest in the programme to be evaluated. Establishing a steering group for the evaluation might particularly be considered when the programme to be evaluated is of major budgetary significance, of controversial nature or when the focus of the evaluation is not only confined to the implementation of the programme but also involves an analysis of its effectiveness and future relevance. Advantages of establishing a steering group include that it:

- ◆ encourages active involvement in the evaluation of the various stakeholders
- ◆ reduces the chances that programme managers will become too closely associated with the evaluator, thus, compromising her/his independence
- ◆ allows for quality control of the evaluation by experts

Creating a steering group, furthermore, helps to ensure that the evaluation is viewed as an inclusive process. Stakeholders are then more likely to have confidence in the evaluation's conclusions and recommendations, especially if they have had the opportunity to influence the design of the evaluation. However, it is important to make sure that all members of the steering group are thoroughly familiar with the terms of reference and that the steering group does not become too large. It may then lose its role as a management body and degenerate instead into a negotiation forum, threatening the impartiality of the exercise.

Regardless of whether or not a steering group has been created, the following types of arrangements may be made between the management and the team entrusted with the conduction of the evaluation, in order to arrange effective monitoring and to ensure that the terms of reference are fully satisfied:

- ◆ frequently held meetings with flexible and informal working procedures
- ◆ regular briefings by the evaluator on the progress of the evaluation
- ◆ preparation of interim reports to indicate preliminary findings of the evaluation
- ◆ discussion and validation of preliminary findings with the stakeholders of the evaluation

In spite of a sound management of the evaluation and the implementation of various feedback mechanisms, problems or changes in circumstances may arise once an evaluation is underway that have to be dealt with by the person(s) responsible for the management of the evaluation. These problems may include:

- ◆ Disagreements between the management and the evaluator on some basic aspects of the evaluation design. For example, if the management asks for a judgement of the effectiveness of a programme despite the fact that the data necessary to form such a judgement are not available.
- ◆ Changes of the evaluation design suggested by the evaluator. If the evaluator, for example, discovers that the original evaluation design cannot be fully carried out within the time required or she/he suggests to change the original design so that more time can be allocated to examining features of the programme which were not part of the original design.
- ◆ Resistance from programme administrators, beneficiaries or other stakeholders encountered by the evaluator, if they, for example, refuse to make data available.

The management of the evaluation needs to be aware of the potential for such problems to arise during the evaluation procedure.





4 CONDUCTING AN EVALUATION

Once all preparations and arrangements are made, the evaluation process can be started. In most cases an external evaluator or a team of external evaluators will be assigned for conducting the evaluation. However, in order to be able to effectually guide the evaluation, to know what can and what cannot be expected from the evaluation, and to be able to assess the quality and make proper use of the evaluation results, it is important for the commissioner as well as for the management of the evaluation to know, what the basic tasks of an evaluation are and how they may be performed. Table 1 summarises the three relevant fields of investigation when evaluating programmes targeted at the promotion of female entrepreneurship.

Table 1: Field of investigation when evaluating programmes promoting female entrepreneurship

	purpose	tasks
evaluating the context and intervention logic of a programme	<ul style="list-style-type: none"> to examine how the programme functions within its economic, social and political context to clarify objectives when these are not clearly expressed to understand how the programme works and how its features hold together to judge the relevance of the programme to judge the coherence of the programme to establish a basis for analysing the programme's effects 	<ul style="list-style-type: none"> to analyse the needs of the target group(s) and the framework conditions of the programme to analyse the hierarchy of objectives and the underlying strategies or assumptions of the programme to compare the needs of the target group with the objectives of the programme to assess the degree of complementarity within a programme and in relation to other similar programmes
monitoring and evaluating the implementation of a programme	<ul style="list-style-type: none"> to provide continuous feedback on the implementation of the programme and identify potentials for improvement to determine whether the programme remains on the right track towards achieving its objectives to identify reasons for successful or unsuccessful performance of the programme to provide evidence for interpreting results and impacts of the programme 	<ul style="list-style-type: none"> to elaborate a set of indicators for continuously monitoring the progress of the programme to analyse how the programme operates, which functions it carries out and how the services are delivered to identify the programme's strengths and weaknesses
evaluating the effects of a programme	<ul style="list-style-type: none"> to determine the effects (results and impacts) of the programme to analyse whether the programme has reached its goals and objectives to judge the effectiveness of the programme to judge the efficiency of the programme to provide evidence for the continued funding or expansion of the programme 	<ul style="list-style-type: none"> to choose a suitable evaluation design to select appropriate indicators and according data collection methods to analyse data and estimate the net effects of the programme to relate the effects obtained to the expected objectives of the programme to relate the effects obtained to the resources mobilised by the programme

The relative importance of each of these fields of investigation strongly depends on the type and purpose of the evaluation to be conducted (see section 3.1.1). *Formative evaluation*, for instance, generally emphasises the assessment of the context, the intervention logic as well as the implementation (and delivery mechanisms) of a programme and – depending on the state of its advancement - is less concerned with effects. Whereas *summative evaluation* stronger concentrates on - but is not restricted to - examining a programme's effects, estimating the degree to which the objectives are achieved, and determining the efficiency of a programme.



4.1 Evaluating the Context and Intervention Logic

Evaluating the context and intervention logic of a programme or measure might be a starting point in each evaluation. Context evaluation examines how a programme functions within the economic, social and political environment of its community and institutional setting, while a programme's logic is defined as the connection between all its elements, i.e. the needs to be met, the strategies chosen, the objectives set (global, specific and operational), the resources mobilised, and the expected effects (results and impacts) of the programme. An analysis of the programme's context and intervention logic provides an indispensable framework for the study of results and impacts and helps stakeholders to understand how the features of the programme are linked to each other and how they relate to a larger policy context.

In an ex ante evaluation, this type of analysis actually constitutes the main bulk of work, as it produces a judgement on the validity and coherence of the programme and its strategies. In an intermediate evaluation, it leads to recommendations for improving the implementation of the programme and its management methods, while in an ex post evaluation this type of analysis is used to clarify objectives when these were not clearly expressed, to structure the collection of information, and to analyse effects more effectively. The analysis of the context and intervention logic shall therefore constitute a key element in each evaluation of programmes or measures targeted at the promotion of female entrepreneurship.

4.1.1 Analysing the Context of a Programme

The aim of context evaluation is to better understand the circumstances and the socio-economic environment in which a programme operates. This involves assessing the needs and assets of the programme's target group(s), examining the framework conditions of the programme (e.g. policy framework, laws, etc.) and exploring the performance of similar programmes by reviewing research literature and documents.

Assessing the needs of a programme's target group(s)

Programmes or measures promoting female entrepreneurship are always conceived with a given set of needs in mind. These needs are the socio-economic problems which the programme seeks to address, expressed from the point of view of its particular target population, i.e. its intended beneficiaries. Assessing these needs is important in order to verify whether the socio-economic problems, that were used to justify the programme, are genuine. In regard to programmes promoting female entrepreneurship, an assessment of needs might, for example, focus on the specific business problems faced by female entrepreneurs (e.g. raising finance) or on their access to conventional support services. When conducted in the scope of an ex-ante evaluation, i.e. during the planning phase of a programme, an assessment of needs may help to ensure that relevant activities or services are implemented, to customise the programme's design to meet the specific needs of female entrepreneurs (e.g. concerning the content and delivery mechanisms of support services) and to justify the programme to potential sources of funding. Determining the needs of the target group can, thus, help to plan relevant and effective interventions or to amend established ones that are no longer relevant. Needs also constitute the judgement reference when assessing the relevance of a programme (see section 4.1.3).

An assessment of needs may be performed by conducting interviews with female entrepreneurs, by informally surveying support services personnel specialised in consulting women entrepreneurs or other key actors in this field, or by simply relying on existing literature.



Analysing the framework conditions of a programme

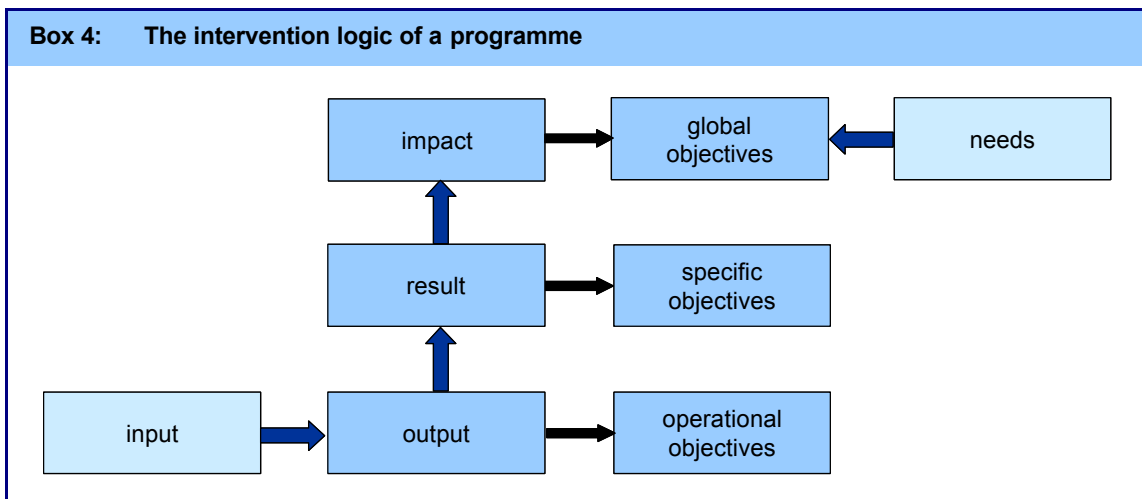
Understanding the circumstances in which a programme operates does not only involve an analysis of the needs of its target group(s) but also of the wider context of the intervention, including the policy considerations to introduce the programme, the legal framework of the intervention, the efforts of related organisations in the relevant field, the relevant economic and financial factors, the impact of social factors, etc. Understanding these environmental influences is required to design a context-sensitive evaluation and helps to interpret findings accurately.

Examining and assessing similar programmes

Exploring the performance of programmes similar to the one to be evaluated may provide information on how, why and in which context other programmes and activities in the realm of female entrepreneurship have been useful. This might best be done by means of a literature review, which may also build the necessary basis to develop a theory as to how the programme will successfully meet the needs of the target group(s) and to claim that the programme will lead to the desired outcomes.

4.1.2 Analysing the Intervention Logic of a Programme

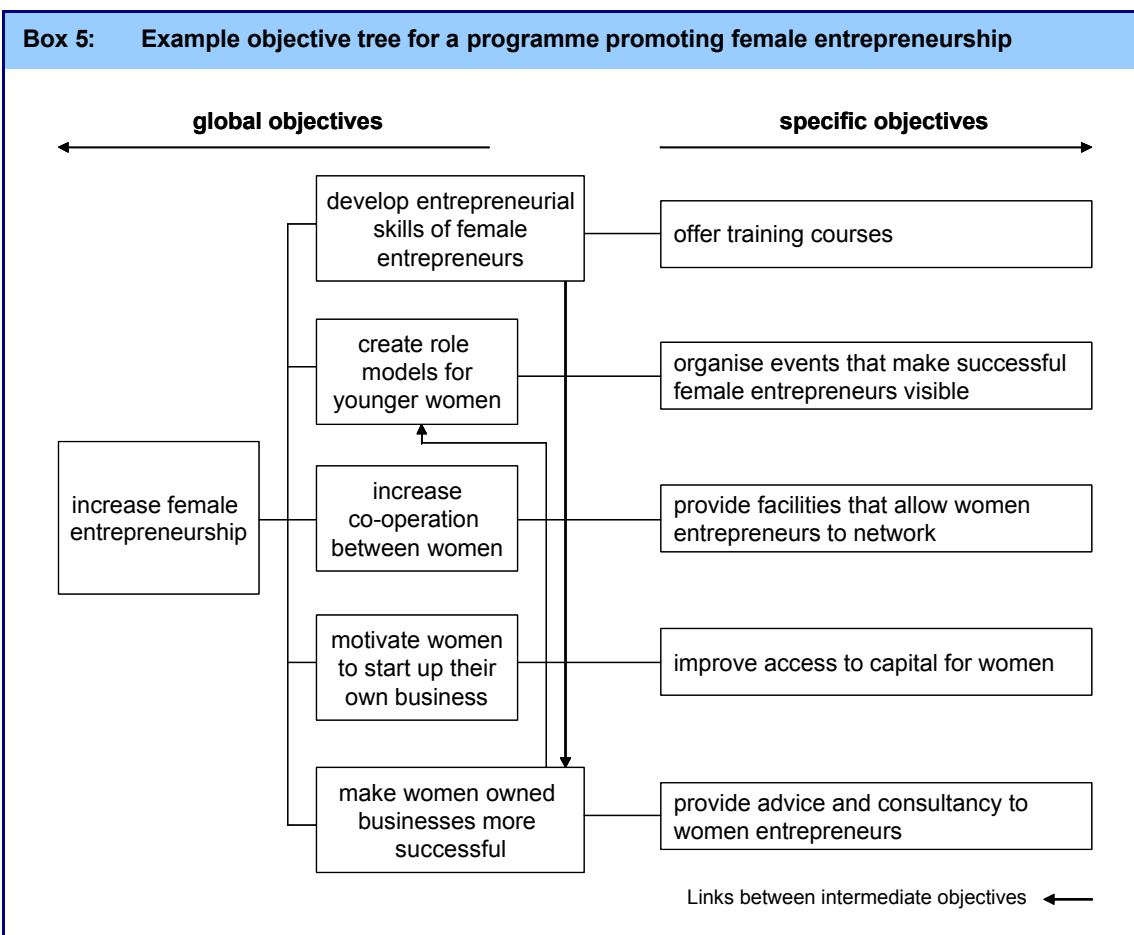
Examining the intervention logic of a programme may help to explain why it has been implemented in the way it has and why certain outcomes have been achieved and others have not. It comprises an assessment of the programme's (global, specific and operational) objectives, i.e. of what impact the operators of a programme want to achieve, as well as of the assumptions or hypothesis used to explain that a given output is going to produce the intended impact. Ideally, the objectives of a programme promoting female entrepreneurship are derived from the needs identified in the scope of the assessment of needs (see section 4.1.1) and are linked to the outputs, results and impacts of the respective programme (see Box 4).



Programme **inputs** are defined as financial or administrative resources mobilised for the implementation of the programme (e.g. available budget for training courses). They are linked to its **outputs**, which subsequently subscribe the product of the intervention (e.g. hours of training financed) and relate to its operational objectives. Programme outputs, in turn, shall lead to **results**, which refer to the most immediate effects of the programme on its direct beneficiaries (e.g. qualifications earned by participants) linked to the specific objectives of a programme, and subsequently to **impacts**, defining the long-term, sustainable effects of a programme (e.g. better management and higher survival rates) and responding to its global objective(s). In Box 4 this intervention logic is described showing that the global objective of a programme, in highly effective programmes, meets the previously identified needs of its beneficiaries.

Analysing the hierarchy of objectives

An important element in evaluating the intervention logic of a programme is the analysis of its *hierarchy of objectives*, which can schematically be represented in the form of an *objective tree*, the global objective being the trunk and the more specific objectives being the branches (see Box 5). The **global objective** (e.g. increasing female entrepreneurship) corresponds to the general aim of the intervention and has to be translated into more **specific objectives** (e.g. to develop entrepreneurial skills among female entrepreneurs) and ultimately into **operational objectives** (e.g. to provide training courses for female entrepreneurs).



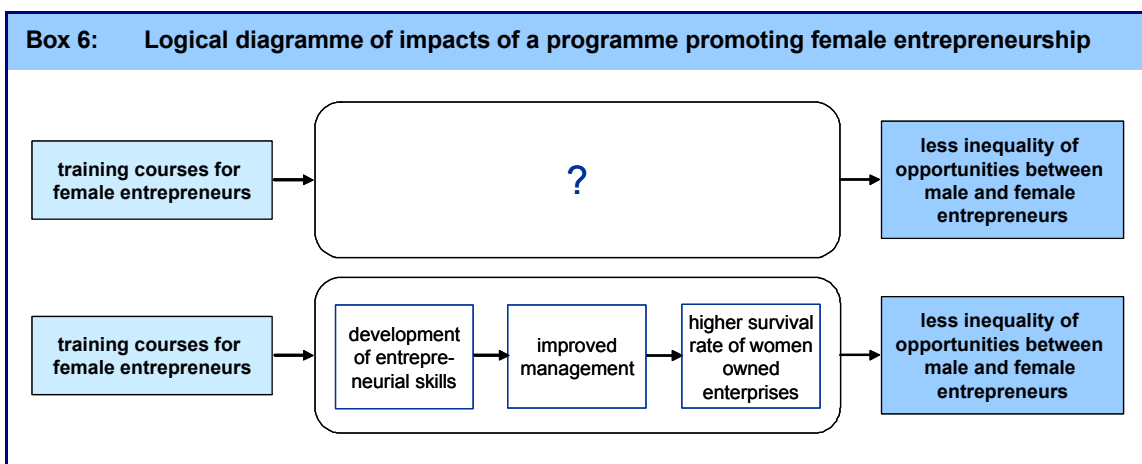
The reason why it is necessary to analyse the hierarchy of objectives is that programme objectives are often of a quite general nature. How they are to be translated into more specific objectives and even more into operational ones, which specify the outputs to be produced, often remains unclear. Hence, it is of crucial importance when evaluating programmes promoting female entrepreneurship to analyse whether the intended outputs (operational objectives) can plausibly be transformed into the expected results (specific objectives), and whether these may lead to the expected impact (corresponding to the global objective).



Analysing the strategy of a programme

When analysing the intervention logic of a programme it is not only important to examine what impact the programme operators want to achieve but also which process is supposed to lead to this impact. In other words, the analysis of the intervention logic should also involve scrutinising the programme's impact hypotheses, i.e. the assumptions on the causal links between the outputs, results and the longer-term impacts of the programme.

Opening the black box of implicit hypotheses helps the evaluator to judge the *validity* of the programme's hierarchy of objectives and hence to answer the question whether the objectives follow logically from the more global objectives to the operational ones and, vice versa, whether the objectives on the lower level constitute a logical contribution to those on the higher level. In the context of female entrepreneurship this means, that it is not only important to clarify which goals a certain programme is expected to achieve, but also which assumptions have been used as a basis for the design of the programme as demonstrated in Box 6.



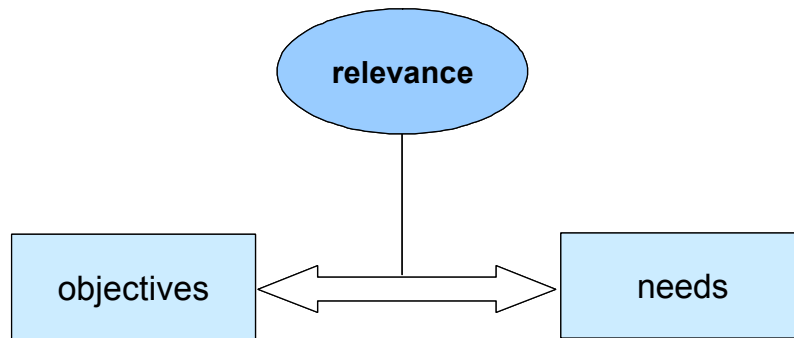
A detailed **logic model** allows stakeholders to clarify the programme's strategies and can also strengthen claims of causality by providing a basis for estimating the programme's effects on endpoints that are not directly measured but are linked in a causal chain supported by prior research. The process of developing a logic model is an interactive one that requires stakeholders to work together in order to clarify the underlying rationale of the programme and the conditions under which success is most likely to be achieved. Gaps in activities, expected outcomes and theoretical assumptions can be identified and the clarity of thinking that occurs from the process of building the model itself provides a starting point for judging the relevance and the coherence of the programme to be evaluated (see sections 4.1.3 and 4.1.4).

The analysis of the intervention logic of a programme usually starts with analysing the official programme documents and, if programme managers agree, may be supplemented by a series of interviews and/or meetings in which participants explain their understanding of the programme's objectives, strategies and expected effects. The evaluation team then has to produce, in a participative manner, a "reconstructed logic" which may serve a basis for further analysis and may be adapted as the activities develop. In fact, an effective logic model will be refined and changed many times throughout the evaluation process as programme staff and stakeholders learn more about the intervention, its operationalisation and how and why it works.

4.1.3 Judging the Relevance of a Programme

The relevance of a programme is one of the main evaluation criterion (see section 3.1.3) and refers to the relation between the programme's objectives and the needs that have to be met, i.e. to the appropriateness of programme's objectives with regard to the socio-economic problems the intervention is meant to solve (see also Box 7). The issue of relevance is closely related to the analysis of the intervention logic of a programme and is particularly important in ex ante evaluation as the focus lies on the strategy chosen. In intermediate and ex post evaluation it is advisable to check whether the socio-economic context and, subsequently, the needs of the target group(s) have changed and, therefore, whether the programme continues to make sense.

Box 7: Definition of the relevance of a programme



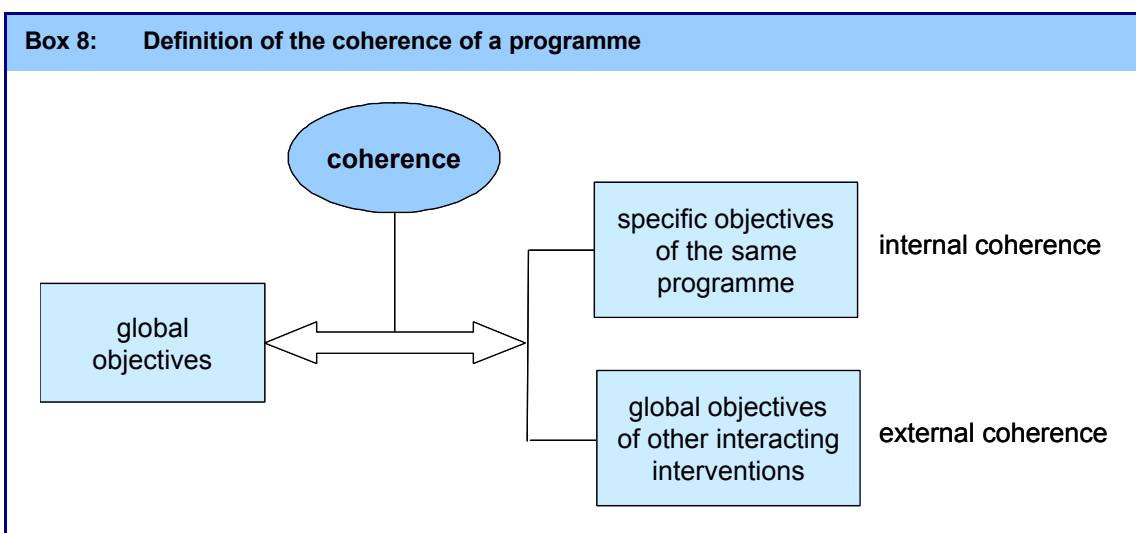
In regard to the domain of female entrepreneurship the main question when analysing the relevance of a programme might be:

Do the programme's objectives reflect the socio-economic problems and needs of (potential) female entrepreneurs in the target region?

An answer to this question may be easily derived from the context analysis in conjunction with the analysis of the hierarchy of objectives of the programme.

4.1.4 Judging the Coherence of a Programme

The coherence of a programme is another main evaluation criterion (see section 3.1.3) linked to the programme's intervention logic and referring to the degree of complementarity or synergy to be found within the programme and in relation to other similar interventions (see also Box 8).



The **internal coherence** refers to the correspondence between the programme's global, specific and operational objectives. It may be judged by analysing the hierarchy of objectives as outlined in section 4.1.2. The corresponding evaluation question might be:

Does the programme have logical objectives, with the specific objectives contributing towards reaching the global ones?

The **external coherence** refers to the correspondence between the programme's objectives and those of other interventions that interact with the programme to be evaluated. External coherence is assured when, for example, a national policy promoting female entrepreneurship and a respective action on European level are implemented in a complementary manner (functional and/or territorial) and contribute towards achieving the same global objective. In the domain of female entrepreneurship the judgement of the external coherence of a programme may correspond to asking the question:

Are the objectives of the programme consistent with the policies and priorities of other institutions or initiatives promoting female entrepreneurship as well as with the strategies of the respective region or country for promoting equal opportunities and/or entrepreneurship?

Similar to the issue of relevance, the analysis of the programme's coherence is particularly important in ex ante evaluation. However, in intermediate and ex post evaluation the first step of the assessment should also be devoted to the programme's logic, as it may happen, for example, that the objectives of the programme were amended midway because there has been no ex ante evaluation, because the ex ante evaluation was not performed correctly or because its conclusions were not taken into account. In that case it is necessary to present a new logic of the programme at the beginning of the evaluation and to check whether the overall coherence of the programme has been maintained despite these changes.

4.2 Analysing the Implementation of a Programme

In order to fully understand how a programme works and what its strengths and weaknesses are, it is of crucial importance to analyse the implementation of the programme by continuously following up and investigating the delivery process of its various measures and activities. The analysis of the implementation of a programme has two dimensions: *programme monitoring*, which refers to the regular collection of (mostly quantitative) data on the progress made by the programme and usually happens under the responsibility of the programme's managers or operators, and *process evaluation* comprising a more qualitative assessment of programme implementation and usually conducted by an (internal or external) team of evaluators. Both dimensions of the analysis of the implementation of a programme are described below.

4.2.1 Monitoring the Implementation of a Programme

Programme monitoring refers to ongoing data collection about a programme's activities and outputs for the purpose of reviewing its progress towards achieving the expected results and assessing its performance over time (see also section 2.1). As a tool for programme managers, monitoring is a mechanism for generating statistics that describe how a programme is evolving and provides assurance that an action or measure is being implemented and provides services as designed. Furthermore, monitoring represents a key information source for process evaluation (see section 4.2.2). Thus, long before an impact evaluation (see section 4.3) may be completed, monitoring can provide on-going feedback to managers on the attainment of service delivery or operational objectives, allowing for adjustments and improvements to be made as found necessary. As such, monitoring serves as a form of an early warning system, alerting programme managers to operational weaknesses that require corrective action. Basically, establishing a monitoring system amounts to asking the question:

Who is doing what, when, where, how often, and with which resources?

A monitoring system may, however, also be developed for assessing service quality by regularly collecting data on customer satisfaction, for example, or may be established with the view to providing data for analysing the cost-efficiency of a programme by assessing the relationship between the resources (inputs) used and the outputs (and eventually results) obtained (see section 4.3.6 for a more detailed explanation of different efficiency indicators). Thus, a variety of specific questions regarding the status and productivity of a programme promoting female entrepreneurship may be addressed by a monitoring system, such as:

- ◆ *Which amount of resources has been spent for which types of activities?*
- ◆ *How many (potential) female entrepreneurs have participated in or profited from the programme?*
- ◆ *Have the planned outputs of the programme been achieved? And have they been achieved within the planned time frame?*
- ◆ *Were the participants of the programme satisfied with the services received?*

Developing an effective and efficient monitoring system requires careful planning and consideration. To be able to monitor a programme's implementation over time and to judge its performance against the objectives set, it is necessary to elaborate a set of respective indicators, to design formats for collecting data on these indicators, and to develop procedures for periodically tabulating and analysing these data. Basically, four levels of indicators may be distinguished:

Input indicator

An input indicator - also often referred to as resource indicator - measures the means (e.g. budget, staff, legal and organisational resources, etc.) mobilised to implement a programme.



Output indicator

An output indicator measures the direct product or service of the respective intervention (e.g. the number of training courses financed for women), i.e. everything that has been obtained in exchange for the input.

Result indicator

A result indicator measures the most immediate impact for the direct beneficiaries of a programme (e.g. the number of women completing the training course).

Impact indicator

An impact indicator measures the consequences of a programme beyond its direct and immediate interaction with its beneficiaries (e.g. the survival rate of enterprises run by women who participated in the training course).

Depending on the state of advancement of a programme, the use of result and impact indicators will be more or less relevant for monitoring. Some results, and even more so impacts, may only be identifiable after several years of implementation and may, thus, be subject to an impact evaluation (see section 4.3). However, an early definition of indicators might help programme managers and stakeholders to assess the intervention throughout its whole life span and facilitate the evaluation of the programme's impacts at a later stage. Moreover, previous thorough reflection on the indicators to be used at the different levels of input, output, result and impact can provide useful insight as to the relevance of the objectives and the intervention logic of a programme (see section 4.1.2). Table 2 provides some examples of output, result and impact indicators for monitoring (as well as evaluating) the progress of programmes targeted at the promotion of female entrepreneurship in the various domains over time.

The final set of indicators to be chosen for monitoring the implementation of a programme should reflect the key activities defined as central to the programme and its logic model (see section 4.1.2) and should be limited to a few main indicators of priority outcomes. At the same time, the indicators should be constructed to reflect the information needs of the numerous stakeholders, including government agents, sponsors, programme managers, and programme staff. Still, the accessibility of data and the cost of allocation must be considered when selecting monitoring indicators, bearing in mind that regular measurement is important so that the system can provide the information in time to make shifts in programme operations possible and to capture changes over time. Routine quality control procedures should also be established in order to check on data entry accuracy and missing information. Thus, in selecting monitoring indicators, programme managers need to consider the following aspects:

- ◆ *The relevance of potential activities to the objectives of the programme:* The indicators shall reflect the programme's strategy and activities identified in mission statements or other relevant programme documents and cover the full range of identified objectives.
- ◆ *The comprehensiveness of the set of indicators:* The set of indicators shall cover inputs, outputs and ideally also service quality (customer feedback) and other results.
- ◆ *The control over the factors measured with the help of indicators:* The programme managers should have control over the outputs or results measured by the indicators.
- ◆ *The validity of the indicators:* The proposed indicators shall reflect the range of outputs and results the programme intends to affect. The data should be free of reporting bias.
- ◆ *The reliability and accuracy of the indicators:* The indicators should ideally be operationally defined in a straightforward manner so that data can be collected consistently over time.
- ◆ *The feasibility of collecting the data:* The effort and money required to generate each indicator shall not be too high unless it is a critically important one.



Table 2: Examples of indicators for monitoring and evaluating programmes promoting female entrepreneurship in the various domains

	output	result	impact
start-ups	number of supported business start-ups by women	percentage of business start-ups by women in high tech sectors	gross value added in newly established businesses after one year
	number of one-stop shops established for potential female entrepreneurs	number of businesses started up by women following assistance	survival rate of assisted business start-ups by women after three years
training	number of supported IT training courses for potential female entrepreneurs	number of women who have attended the courses and improved their IT skills	number of business start-ups in the IT branch by female entrepreneurs
	number of training courses financed for female entrepreneurs	number of female entrepreneurs who have gained a qualification	survival rate of enterprises run by women who participated in the training
advice and consultancy	number of hours of consultancy services financed for female entrepreneurs	rate of satisfaction among female entrepreneurs who received consultancy services	percentage of female entrepreneurs who have increased their activities due to consultancy
	number of female entrepreneurs who received advice in the field of product marketing	percentage of female entrepreneurs who increased their marketing activities due to advice	percentage of increased turnover due to improved marketing activities in these enterprises
mentoring	number of paid or volunteer mentors with appropriate skills	number of successful matches between mentor and mentee	number of women starting up their own business due to mentoring service
	frequency and lengths of meetings between mentor and mentee offered	rate of satisfaction with the mentoring relationship among mentees	percentage of women with an improved work/life balance due to mentoring
information	number of female entrepreneurs provided with information on export markets	number of female entrepreneurs becoming new exporters	percentage of export sales in the turnover of assisted enterprises after 18 months
	number of inquiries related to research and development issues	number of female entrepreneurs doing research and development	number of patents applied for or provided to female entrepreneurs
funding	number of female entrepreneurs who received grants for investment projects	number of investment projects realised successfully by female entrepreneurs	annual value added generated by investments of female entrepreneurs after one year
	number of guarantees provided for business take-overs by women	number of enterprises successfully taken over by women	survival rate of enterprises taken over by women after 3 years
networking	number of conferences and exhibitions organised for female entrepreneurs	average number of newly established contacts by participating female entrepreneurs	number of long-term commercial business relationships established by female entrepreneurs
	number of women visiting the web page of a networking service	number of women participating in online discussions	number of women regularly using network technologies for business opportunities



4.2.2 Evaluating the Implementation of a Programme

Evaluating the implementation of a programme amounts to conducting a so called *process evaluation*, which adds a qualitative dimension to the descriptive statistics produced by a monitoring system. Process evaluation focuses on how a programme was implemented and identifies the procedures undertaken as well as the decisions made in developing the programme. It describes the programme's operations, which functions it carries out and how the services are delivered. Like monitoring, process evaluation addresses whether the programme was implemented and is providing services as intended, however, by additionally documenting the programme's development and operation, process evaluation assesses reasons for successful or unsuccessful performance and provides information for potential replication. Generally, process evaluation of programmes or measures targeted at the promotion of female entrepreneurship may contribute to answering the following types of questions:

- ◆ *How is the programme offered? Who uses the programme?*
- ◆ *What problems were encountered in implementing the programme? How were they resolved?*
- ◆ *Is the programme implemented as specified and, if not, how do the operations differ from those initially planned?*
- ◆ *How are the operational responsibilities distributed?*
- ◆ *What is the view of the programme from the perspectives of the programme's staff, participants and other stakeholders?*
- ◆ *What lessons might be learned from the implementation of the programme that might be useful for its future modification?*

The focus of a process evaluation varies depending on the phase of the programme and the overall purpose of the evaluation. When evaluating a *new programme* that is still in its formative stage, process evaluation serves the purpose of assisting decision makers in documenting the programme's evolution and assessing whether modifications are linked to the objectives, the relevant contextual factors, and the needs of the programme's target population. In this regard, the evaluator must understand the functioning and development of the programme. Typical evaluation questions in this context are:

- ◆ *To what extent does the programme look and act like the one originally planned?*
- ◆ *Which initial strategies or activities of the programme are being implemented? Which ones are not? Why or why not?*
- ◆ *How is the programme working and what additional changes may be necessary?*
- ◆ *Does the programme reach the addressees it intends to serve?*

In case of an *established programme* that has been running for several years, the evaluation of its implementation is usually designed as a continuous evaluation, feedback and improvement loop. This type of continuous or on-going process evaluation provides feedback and helps to recognise which activities work well and which ones need modification. Examples of questions addressed in this type of evaluation include:

- ◆ *Which operations of the programme work well? Which ones do not? Why or why not?*
- ◆ *Which strategies have been successful in encouraging client participation and involvement? Which ones have been unsuccessful?*
- ◆ *Which components are the most important ones to the programme's success?*
- ◆ *How effective is the organisational structure in supporting the implementation of the programme? What changes need to be made?*



Summing up, process evaluation focuses on how a programme is implemented in order to determine whether it has remained on the right track towards the achievement of its objectives and to identify influencing factors that may have prevented the attainment of certain objectives or goals. This might involve a qualitative description of how the core activities were implemented, of the modalities of the programme's management, of the distribution of responsibilities, and of key factors influencing the outputs as well as the results of the programme. In this respect, it might be important to know, that the following issues have, among others, proven to be particularly crucial to the successful implementation of support programmes targeted at the promotion of (female) entrepreneurship and should therefore receive particular attention when evaluating the implementation of programmes or measures in this domain:

◆ **The visibility of the programme**

Is the programme well communicated and well known to its potential clients? Is it being effectively promoted? Are (potential) female entrepreneurs in the target region aware of the existence of the programme?

◆ **Convenient access to the programme**

Is the programme or service easily accessible for its potential clients (e.g. regarding its geographical location, its communication procedures, administrative requirements, etc.)?

◆ **The supply of a distinct package of services**

Does the programme offer a distinct package of services required by the client? Does it take account of the differing needs of (potential) female entrepreneurs (e.g. focused short-term help vs. intensive start-up assistance or individual vs. group-based support)?

◆ **The client-orientation of the programme**

Are the nature and form of the programme as well as its delivery mechanisms suitable to the needs of (potential) female entrepreneurs (e.g. in terms of scheduling and length)?

◆ **The assurances of the quality of the programme**

Is a gender aware assessment of needs carried out in order to find out what the specific needs of (potential) female entrepreneurs and the barriers affecting their economic participation are? Does the programme apply quality assurance systems (e.g. with reference to general or self defined quality standards)? Do regular follow-ups with the clients take place?

While a variety of research methods are available for conducting process evaluation, the most commonly chosen approaches are based on qualitative methods of data collection (e.g. semi-structured interviews with those involved in the implementation of the programme, focus groups with the programme's staff or participants, case studies, or document reviews) in order to obtain an in-depth picture of the programme's development and the environment in which it operates (see section 4.3.3 for overview of the different data collection techniques). Ideally, these qualitative data are combined with quantitative ones (e.g. on the resources used, the activities offered or the initial results achieved) obtained from the monitoring system of the programme (see section 4.2.1).

Evaluating the implementation of a programme is a vital source of information not only for the analysis of how a programme works, but also for interpreting results. It increases the relevance of impact evaluation (see section 4.3), as knowing why a programme achieves its objectives is more important than just knowing that it does. In fact, without knowing what was implemented and why, it is virtually impossible to select valid effectiveness measures or to show causal linkages between the programme's activities and its outcomes. In this respect the evaluator should also carefully refer to the logic model of the programme (see section 4.1.2), when analysing the programme's operations, although it is permissible and important in process evaluation to revise the original logic model in light of the findings obtained during the evaluation.



4.3 Evaluating the Effects of a Programme

Examining the effects of a programme amounts to carrying out an *impact evaluation*, which analyses whether a programme has successfully achieved its goals and objectives and which changes are brought about by the programme. While *monitoring* and *process evaluation* focus on whether an action or measure was implemented and carried out as designed (see section 4.2), impact evaluation addresses whether the programme's activities had their intended effects, namely their expected results as initial effect and their intended impacts as longer-term effect.

Examining the effects of a programme may take place before the approval of an intervention (ex ante evaluation), after completion of a programme (ex post evaluation), or at any stage in between (intermediate evaluation). Ex ante impact evaluation forecasts potential results and impacts as part of the planning, design and approval of a programme. Ex post or intermediate impact evaluation is critical to identifying actual results and impacts, enabling to judge the effectiveness and efficiency of a programme, and to providing evidence for its continued funding or expansion/replication. Sample evaluation questions that may relate to an ex post or intermediate impact evaluation of a programme targeted at the promotion of female entrepreneurship include:

- ◆ *What has happened as a result of the programme?*
- ◆ *What are the long term impacts of the programme?*
- ◆ *What are the unintended effects of the programme?*

Given the nature of impact evaluation, it typically involves more sophisticated and complex methodological approaches than monitoring and process evaluation. In order to determine whether an intervention actually had an impact requires a consideration of what would have happened or would be expected in the absence of the programme. As a result, impact evaluation requires the establishment of a comparative research design, the selection of appropriate indicators, the collection of corresponding data, and the calculation of net effects that can be directly attributed to the respective action. These issues are dealt with in the following sections.

4.3.1 Choosing a Suitable Evaluation Design

Impact evaluation must be carefully designed if it is to produce valuable results. An evaluation design is a model which is used to provide evidence on whether observed effects are indeed attributable to the programme to be evaluated. An evaluation design that is not carefully constructed can mask inherent biases and values, waste valuable resources by gathering data that do not address important evaluation questions or lead to an inaccurate or invalid interpretation of data. The evaluation questions, along with the ultimate scope and purpose of the evaluation (see section 3.1) are critical to determining an effective evaluation design. The choice for a specific design allows evaluators to build a framework for testing hypotheses about a programme's effects (see section 4.1.2) and to choose the appropriate techniques for collecting the necessary data and information. A classification of different evaluation designs includes experimental, quasi-experimental and non-experimental designs. In the following, the key elements of these three types of design are illustrated. Their strengths and limitations as well as their general requirements in terms of resources are summarised in Table 3.



Table 3: Evaluation designs for estimating net effects of a programme

	key elements	advantages	limitations
non-experimental design a) before and after comparison b) cross-sectional comparison c) panel design	examines only changes for participants of a programme, no comparison with non participants	comparatively easy and inexpensive to conduct	cannot control for the effects that would have occurred without the programme the extent to which results can be applied to other groups or other settings is limited
quasi-experimental design	compares effects for participants of a programme to effects for a "comparison group" that did not participate, but is similar to the group of participants in its main characteristics	allows for a direct attribution of effects to the programme gives a reliable estimation of the net effect of a programme	there is a risk that unrelated events may affect outcomes collecting comparable data on comparison groups can be difficult
experimental design	compares effects for participants of a programme to effects for a "control group" that has been selected randomly	is highly valid and reliable has strong causal inference and is statistically convincing	random assignment evaluations are not always the ideal choice in real life settings and can be quite expensive some interventions are impossible to study through randomised experiments

Non-experimental evaluation designs

Non-experimental designs only examine changes for participants of a programme and do not compare the effects for participants to individuals or groups who did not participate in the respective intervention. The evaluator can choose a) to compare participants before and after they receive a service, b) to make cross-sectional comparisons after a service has been delivered, or c) to conduct a panel study. In a *before and after comparison*, effects for groups of participants that entered the programme at a specific time are measured before and after an intervention. The assessment of the programme's impact is inferred from the differences in the average score for the group before and after participation. This simple design is often used to assess whether knowledge, attitudes, or behaviour of the group changed after exposure to an intervention. *Cross-sectional comparisons* are based on surveys of a programme's participants conducted after the intervention is completed. Data collected with this design can be used to estimate correlations between the effects experienced by individuals and differences in the duration, type and intensity of the services they received. This allows drawing some conclusions about plausible links between the effects and the different types of services within the programme. However, definitive conclusions about what caused what cannot be drawn, as there is no comparison group that would let you say "it happened for those who got services, but not for those who did not get services." Finally, *panel designs* use repeated measures of the outcome variables for individual participants in an intervention. In this design, effects are measured for the same group of participants, often starting at the time they enter the programme and continuing at intervals over time.



Quasi-experimental evaluation designs

Quasi-experimental designs compare effects for participants of a programme to effects for a *comparison group* that did not participate in any of the programme's activities. The critical difference between quasi-experimental and experimental designs (see below) is that the decision on who participates in the programme is not random in quasi-experimental designs. Comparison groups are made up of individuals as similar as possible to the participants of the programme on factors that could affect the selected outcomes one wants to measure. Statistical techniques are then used to control for remaining differences between the groups. The most common technique is called 'matching pairs', where the evaluator constructs a comparison group by matching individuals who do not participate in the programme to individuals who do participate on a selected set of characteristics.

Experimental evaluation designs

Experimental designs are considered the "gold standard" in evaluation. Experiments require that individuals or groups are assigned at random to one or more groups prior to the start of the programme's activities. A *programme group* receives services designed to achieve clearly specified outcomes, whereby a *control group* receives no such support. Because chance alone determines who receives the programme's services, the groups can be assumed to be similar in all characteristics that might affect the outcome measures. Any difference between programme and control group, therefore, can be attributed with confidence to the effects of the programme. Although experiments are the preferred design for an impact evaluation on scientific grounds, random assignment evaluations are not always the ideal choice in real life settings. Some interventions are inherently impossible to study through randomized experiments for legal, ethical or practical reasons. Also experiments typically require high levels of resources, money, time, expertise, and support from programme staff, government agencies, sponsors, etc. Those conducting an evaluation as well as those commissioning it have to ask themselves, whether the results that are likely to be obtained from this approach justify the investment.

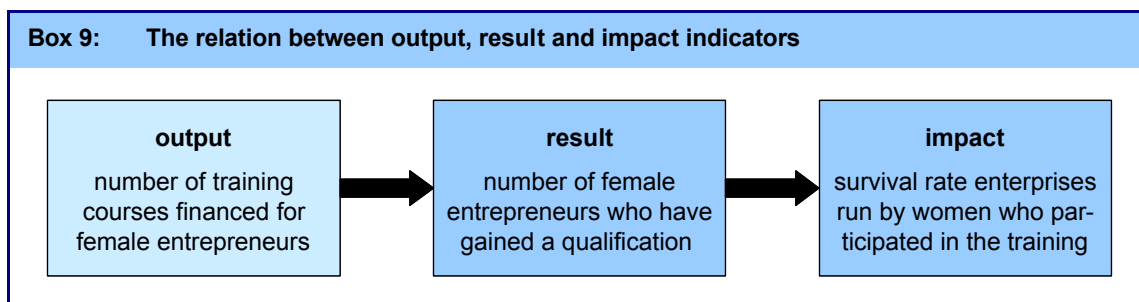
4.3.2 Selecting Appropriate Indicators

An important step in any impact evaluation after deciding on a respective research design is to – possibly with the help of programme staff and key stakeholders – think through the possible outcomes of the programme to be evaluated. In order to determine how effective a programme is, one needs to have some idea on how well outcomes are being achieved and on how, i.e. with which indicator, to measure them. The concept of indicators has already been introduced in section 4.2.1. While monitoring indicators assess the progress of a programme towards the achievement of its objectives, indicators for impact evaluation may be described as a measurable approximation of the effects one is attempting to achieve. They translate general concepts regarding the intervention and its expected effects into specific measures that can be interpreted and that usually address criteria that will be used to judge the intervention.

Most important for assessing a programme's effects are result and impact indicators. While *result indicators* measure the most immediate impact for the direct addressee of a programme, *impact indicators* represent the consequences of the programme beyond its direct and immediate interaction with the addressees, i.e. the socio-economic changes which can be attributed to the programme (see also section 4.2.1 for a definition of the different types of indicators). The indicators to be selected for each outcome depend on the evaluation team's perspective about what is the most accurate measure of the stated outcomes and on the resources available for data collection.



The starting point for selecting appropriate indicators for evaluating the effects of programmes targeted at the promotion of female entrepreneurship are the general evaluation criteria to be applied in the evaluation (see section 3.1.3). After the evaluation criteria have been translated into specific evaluation questions that specifically relate to the programme under investigation (see section 3.1.4), each evaluation question has to be made operational; i.e. measurable by defining corresponding indicators and selecting the according data collection methods (see section 4.3.3 for an overview of different techniques for data collection). Ideally the result and impact indicators selected for evaluating a programme's effects logically relate to the output indicators established for monitoring the programme's progress in order to allow causality analysis as demonstrated in Box 9.



With reference to the causality chain illustrated in Box 9, training for female entrepreneurs leads to higher qualification levels among women participating in the training course and in consequence to a higher survival rate among the enterprises owned by these women. However, there may well be found a growing number of female entrepreneurs taking training courses in the monitoring data, but if these courses do not correspond to the needs of these women, the intervention may not contribute to higher entrepreneurial skill levels among female entrepreneurs at the *result level* and, thus, to a higher survival rate of enterprises run by women at the *impact level*. This would imply that the hidden assumptions of the programme's planners about the causal links between the programme and its supposed effects, i.e. the intervention logic of the programme (see section 4.1.2), cannot be verified. Table 2 in section 4.2.1 provides an example of a set of result and impact indicators for evaluating different types of measures promoting female entrepreneurship interrelating to the respective output indicators.

However, particularly when evaluating programmes targeted at the promotion of female entrepreneurship, *individual client effects* should be determined in addition to the *socio-economic effects* of a programme tackled in Table 2. Often the objectives of a programme may be expressed in terms of service delivery or system objectives only (e.g. reducing inequalities between male and female entrepreneurs), rather than in terms of individual client objectives, stating how clients' lives will improve as a result of the programme (e.g. gaining life and business skills necessary to be self-reliant and economically independent). Yet, when thinking about the purpose of social and human services programmes, it has to be realised that one of the most important outcomes is the effect on individuals/participants. Evaluating individual client effects amounts to answering the question:

What difference does the programme make in the lives of those served?

The types of effects one may attempt to achieve with programmes promoting female entrepreneurship at the individual client level include changes in circumstances, status, quality of life, attitude or behaviour, knowledge, skills, etc. These types of effects should be analysed in the scope of all impact evaluations of programmes targeted at the promotion of female entrepreneurship - although they might not be stated as explicit objectives of the intervention - and should therefore be considered, when elaborating corresponding evaluation indicators. Furthermore, certain programmes often produce outcomes that were not listed as objectives when the programme was launched. Impact evaluation should also attempt to discover these unanticipated or *unexpected effects* by elaborating according indicators.



4.3.3 Applying Suitable Techniques for Data Collection

Programme effects can only be estimated if data are available. The most immediate source of data about a programme should normally be the monitoring system (see section 4.2.1). However, monitoring data will usually be restricted to outputs or some immediate results. Thus, additional data and information has to be collected. Varying techniques from quantitative survey analysis to qualitative case studies suitable for evaluating programmes promoting female entrepreneurship exist. However, there is no single method with the help of which all programmes, actions or measures can be effectively evaluated. The choice of the data collection technique follows from the choice of the evaluation design (see section 4.3.1) and the indicators selected for evaluation (see section 4.3.2), and has to be determined by the nature of the intervention as well as by the purpose and scope of the evaluation at stake (see section 3.1.1). In the following, some of the most frequently used information and data collection methods for evaluation purposes are described, differentiated by techniques for collecting *primary data*, which are taken directly from original sources or collected first hand and *secondary data* that have undergone extensive manipulation and interpretation.

Collecting secondary data

Secondary data have the advantage that they already exist. However, they rarely contain all the information needed and usually cover a geographical area that contains many more people than are served by the programme to be evaluated. Secondary data do usually not allow analysing the effects of a programme without the collection of additional data and, thus, will mostly have to be combined with primary data. Generally, the following techniques for secondary information and data collection are applied as a starting point in each evaluation. They are particularly useful in providing information with regard to the analysis of the context and the intervention logic of a programme (see section 4.1).

Review of programme documents

It shall usually be possible for the evaluator to obtain basic information on the programme being evaluated by reviewing the general programme files, financial and administrative records and specific programme and policy documents. Any gaps in the available secondary data may then be identified and primary data collection methods applied to complete the picture. Programme document reviews provide the evaluator with valuable background information on the intervention and its environment and may hence put the effects of the programme in context. They also produce a useful framework and basis for a subsequent primary data search. The commissioners of an evaluation should make sure that the evaluation team gets access to the necessary programme and management documents. Document reviews tend to be a relatively quick and cost saving data collection technique. However, programme documents typically only shed light on programme outputs but not on results or impacts.

Literature analysis

Another source of secondary data is a literature review, which enables the evaluator to make the best use of previous work in the relevant field and hence to learn from the experience and findings of those who have carried out similar or related work in the past. There are various types of publications that can be analysed in the frame of a literature search, such as published papers, reports and books prepared by academics, experts and official organisations, or specific studies in the area under investigation including past evaluations. A literature review is a relatively inexpensive and efficient way of collecting secondary data. Furthermore, past research on the issue of female entrepreneurship, for example, may suggest hypotheses to be tested, specific techniques for overcoming methodological difficulties, or evaluation issues to be examined in the current evaluation. The weaknesses of a literature review are those associated with the inherent nature of secondary data: findings or data may not be relevant or compatible enough with the evaluation issues to be of use in the current work.



Statistical sources

In some cases, statistical data, such as the number of new businesses created by female entrepreneurs in a specific area, for example, may be useful for evaluation purposes. This type of secondary data may be directly obtained from national public statistic offices or other organisations which produce and publish statistical data. Unlike monitoring data which concentrate on the direct addressees of an intervention, statistics generally also encompass people or businesses in an area that have had no contact with the programme. This information may be relevant within the framework of impact evaluation, although it does not provide an estimated impact, but rather indicates changes that may be due to several causes, including exogenous factors. Statistical data might also provide useful information on the context of a programme.

Collection of primary data

There are many types of strategies and sources for collecting primary data. In the following, the most common types of methods used in the field of evaluation are presented. Thereby, it can be distinguished between quantitative and qualitative methods of data collection. *Quantitative techniques* aim at measuring actual impacts, while *qualitative techniques* are particularly suitable for gathering more in-depth information on processes, experiences and opinions. Also in the course of impact assessment, it is necessary to undertake some degree of qualitative analysis in order to be able to evaluate the causes of impacts which have been observed and to describe more complex types of social impacts difficult to measure. The mix of techniques most appropriate for a specific evaluation will depend on the nature of the intervention to be evaluated as well as on the purpose and scope of the evaluation at stake. However, generally, questionnaire surveys are the most widely used method for collecting quantifiable information, while interviews with operators or programme participants, focus groups with programme managers, case studies or observations are among the most common tools for gathering qualitative information. Table 4 summarises the overall purpose, advantages and challenges of these techniques.

Table 4: Overview of techniques to collect primary data

	overall purpose	advantages	challenges
questionnaire surveys	to estimate and quantify a programme's impacts based on homogenous data	allow direct measure of impacts produce structured and generalisable information	costs of constructing a representative sample can be high
individual interviews	to gather qualitative data and opinions of those affected by a programme	provide more in-depth and detailed information than a questionnaire survey	interviews can be difficult and costly to analyse interviewer can bias clients' responses
focus groups	to explore complex issues in depth by means of group discussions	provide an efficient way to get a wide range of in depth information in short time	specific skills are required to manage group dynamics responses might be difficult to analyse
case studies	to provide a complete illustration of a given situation	provide a detailed analysis of complex processes	well-suited for in-depth understanding of a situation but with limited potential for generalisation
observations	to directly observe activities of programme staff and participants	useful to better understand the programme's activities through first hand observation	people may change their behaviour when they are under observation limited validity



Most evaluations will use a combination of data collection techniques in order to address a wide range of issues and to be able to compensate for the weaknesses associated with one technique by the strengths of another. In the following a brief overview is given on the techniques for collecting primary data applied most frequently in the scope of evaluation.

Questionnaire survey

Questionnaire surveys are a versatile method for collecting primary data. Usually they are carried out by mail, telephone or, more recently, by e-mail in order to collect homogenous data on a programme's impacts by means of close-ended standard questions, i.e. the respondents choose from among pre-defined responses offered in the questionnaire or by the interviewer. The questions asked may be descriptive (e.g. *How did you learn about the existence of the programme?*), normative (e.g. *Does the programme correspond to your needs?*) or causal (e.g. *What would you have done without the programme?*). Questionnaire surveys are used extensively in evaluation and can be applied in the frame of different designs for impact analysis. In experimental and quasi-experimental designs data collected by this means give a reliable estimation of a programme's impact which is the difference between the results of the two surveys. However, impact analysis can also, and with lower costs, be directly carried out among the beneficiaries of a programme, e.g. among (potential) female entrepreneurs, by applying non-experimental designs. When surveying female entrepreneurs, who have participated in an export promotion programme, for example, the questionnaire might include questions such as: *What was the value added for your business generated by the assistance?* or *How many jobs have been created in your enterprise thanks to the programme?* Asking this sort of questions makes the construction of a policy-off situation unnecessary; the surveyed people conceive the policy-off situation when they answer. The weakness of this approach, however, is that beneficiaries tend to over-estimate effects for various reasons. Still, the questionnaire survey has the advantage of producing structured, homogenous and generalisable information. It is therefore particularly useful for quantifying indicators, including monitoring indicators. However, surveys require profound expertise in their design, conduction and interpretation; if survey techniques are misused, the data obtained will be invalid and unreliable.

Individual interviews

Individual interviews are used to gather qualitative information and opinions of those affected by a particular programme, its context, implementation, results or impacts. They can be conducted with programme managers, a limited number of programme participants, or with experts not directly involved with the intervention. Interviews may be used in all phases of the evaluation, but are particularly useful when conducting a context analysis (see section 4.1.1) or a process evaluation (see section 4.2.2). As interviews provide in-depth and detailed information, they can indicate whether a programme was implemented as originally planned, and if not, why and how it has changed. Therefore interviews are often used as part of formative evaluations; they may, however, also be helpful in the scope of summative evaluations of programmes which intend to achieve changes in actor's behaviour or perceptions (e.g. personal development training programmes for female entrepreneurs) as opposed to interventions aiming at more tangible, "harder" outcomes (e.g. programmes supporting enterprise start-ups). The inside knowledge gained from interviews can also provide an in-depth understanding of hard-to-measure concepts, such as empowerment, for example, which might be an important issue in the frame of programmes promoting female entrepreneurship. Interviews carried out with experts in a specific domain may be applied for measurements in areas where objective data are deficient. However, individual interviews only take into account situational and individual factors of those interviewed; it is difficult to draw general conclusions from them. Thus, interviews may provide insight into the mechanisms of programme implementation and allow for an exhaustive identification of results as well as of the causal links peculiar to a programme, but usually cannot measure impacts. To be effective, the interview technique requires a lot of time and the contribution of professionals. Specific skills are needed to plan, conduct and interpret an interview.



Focus groups

A focus group involves several people with the same characteristics and provides qualitative information during a targeted discussion. Focus groups are particularly valuable for analysing themes or fields which give rise to divergent opinions or which involve complex issues that need to be explored in depth. With this technique evaluators seek to understand attitudes through a series of group discussions (normally six to eight participants) guided by one researcher acting as a facilitator, with another researcher present to take detailed notes (discussions may also be tape-recorded). Normally, about five or six general questions are selected to guide open-ended discussions lasting about an hour and a half. By playing on the interaction and confrontation of points of view, this technique serves to reveal the participant's perceptions on a given theme that may concern the relevance of the evaluated programme, its implementation, results or impacts. Focus groups make it possible to bring together, simultaneously or sequentially, several stakeholders of a programme (managers, operators, beneficiaries, etc.) and to collect a large amount of qualitative information in a relatively short space of time. However, specific skills are required for managing the group dynamics and obtaining a balanced discussion while avoiding the dominant influence of opinion leaders, for example.

Case studies

Case studies involve an in-depth analysis of the data collected on specific cases, which may concern individuals, programmes or organisations. They are usually used for evaluations aimed at a detailed analysis of complex implementation processes of a programme and the representation of its impacts. Case studies generally apply to situations that require an exploratory field study due to the lack of available information. The cases must, however, be chosen carefully as they should either be representative of the programme as a whole, or should illustrate a specific point, such as a particularly effective action or an approach which was found to have serious deficiencies and which should therefore be avoided in future, for example. The case study is intended to be the most complete illustration possible of a given situation, so as to give a precise image of current phenomena and to understand their causes. It follows that case studies are based on multiple data sources, including a wide range of informal methods which have less precise procedures and rely to a large extent on subjective judgements, such as field visits or unstructured interviews with key informants of a programme, for example. Case studies are relevant for giving a view of processes and complexities that are impossible to see in any other way. However, it is impossible to generalise, in a statistical sense, the results obtained from one or more selected cases to a global population. Usually case studies are also quite expensive and time consuming to carry out.

Observations

Observations involve the evaluator making on-site visits to locations where the programme to be evaluated is in operation and directly observing the activities of programme staff and participants. The value of observations is that the evaluator can better understand programme activities and effects through observing first hand what is happening and how people are reacting to it. Observations might be especially useful when conducting process evaluation, however, in the course of interaction, a number of seemingly insignificant details may also be important in assessing the programme's impact. Information gathered through observation will allow the evaluator to examine the programme's physical and social setting, its staff's and beneficiary's characteristics, group dynamics, and formal and informal activities. The evaluator may also become aware of aspects of the programme that may not be consciously recognized by participants or staff and learn how programme activities change and evolve over time. However, the internal validity (i.e. the absence of technical bias) and the external validity (i.e. the representativeness of results) may be limited since another person making the same on-site visit may derive different observations to those of the evaluator. In addition, there is the specific problem of the so called "Hawthorne effect"; the fact that people (e.g. programme staff, beneficiaries, etc.) may behave differently from their normal patterns when they know that they are being observed.



4.3.4 Estimating Net Effects of a Programme

After having decided on an evaluation design, selected appropriate indicators and the right data collection methods, and after having collected the respective data, the data have to be analysed with a view to assessing the effects of the programme to be evaluated. Many statistical and non-statistical techniques for analysing and interpreting quantitative data exist (e.g. descriptive statistics, regression, variance or factor analysis, macro-economic models, etc.). Due to their rather technical character, these techniques are, however, not being explained in detail here. The purpose of this section is rather to point to the necessity of assessing the *net effect* of a programme, i.e. the effect really attributable to the intervention, as opposed to the *gross effect*, which is the change observed following a public intervention, as well as to illustrate some specific concepts which are particularly crucial in the course of estimating net effects.

In order to evaluate the net effect of a programme, it is necessary to subtract from the gross effect the changes which would have occurred in the absence of the programme and which are therefore not attributable to the intervention, since they are produced by confounding factors (counterfactual situation). This might be explained with the help of the following example: Assuming that the number of enterprises started up by female entrepreneurs appears to be stable (gross effect equal to zero), but it is estimated that without the support programme 100 enterprises would not have survived (counterfactual situation). Thus, it may be concluded that 100 businesses run by female entrepreneurs were maintained due to the programme, which equals the net effect of the respective intervention.

The example shows, that evaluating net effects amounts to:

- 1) describing what would have happened without the intervention (counterfactual situation)
- 2) comparing the counterfactual situation with what actually happened (gross effect) and estimating the difference

When using a true experimental design in evaluation (see section 4.3.1), the counterfactual situation is represented by the control group. Analysing the data and estimating the net effect in this case is straightforward as any difference between the programme and the control group can be attributed with confidence to the net effect of the intervention. However, most often it is not possible to apply an experimental design. Other possibilities to describe what would have happened without the intervention and, thus, to simulate the counterfactual situation are either to apply a quasi experimental design, i.e. to construct a control group, or to directly measure net effects by surveying the participants of a programme, for example. Still, the latter technique is useless when it comes to measuring effects for other affected parties than the target group(s) of a respective intervention.

Central concepts when evaluating the net effect of an intervention are deadweight, displacement, substitution, additionality and leverage effects. Not all of these concepts might be relevant in the domain of female entrepreneurship, however, for the sake of completeness they are briefly described below:

Deadweight loss

Deadweight is defined as the effect which would have arisen even if the programme had not taken place. For example: a female entrepreneur receives assistance for setting up an enterprise in the IT branch. In the survey she states that the support had enabled her to gain better knowledge on the IT market, but that she would have set up an enterprise in the IT branch anyway, even without receiving support. Thus, there is a certain degree of deadweight as the set up of the enterprise cannot be attributed entirely to the intervention.



Displacement effect

The concept of displacement is used to describe a situation where the effect of an intervention on a particular individual, group or area is only realised at the expense of other individuals, groups or areas. If, for example, female entrepreneurs used the assistance offered within a programme targeted at the increase of business start-ups by women in rural areas to move their premises from the centre to the outskirts of a town, then the increase of women led enterprises in the outskirts resulted in the disappearance of female entrepreneurs in the centre of the town and, thus, in a displacement of businesses run by women from one area to another. When they are not intended, displacement effects must be subtracted from the gross effect in order to obtain the net effect of an intervention.

Substitution effect

The term substitution refers to an effect obtained in favour of a direct addressee of an intervention, but at the expense of a person that does not qualify for the respective action. A substitution effect is a specific type of displacement effect. Assuming, for example, that a woman unemployed for a long time found a job owing to a specific labour market intervention, but the job was obtained only because a male employee was granted early retirement. If the objective of the programme was the redistribution of jobs in favour of women, the effect can be considered positive, if it was to increase active labour market participation, it appears to be negative. Evaluation has to determine, with regard to the objectives of an intervention, whether the substitution effect can be regarded beneficial or not. When it is not beneficial, the substitution effect must be subtracted from the gross effect in order to calculate the net effect of an intervention.

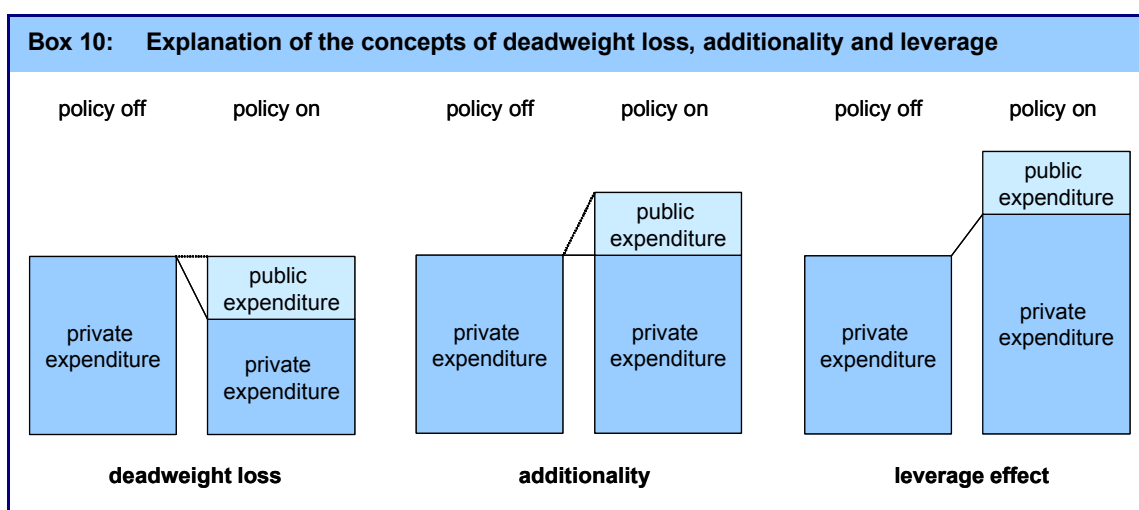
Additionality

Additionality is the converse of deadweight loss (see above); it occurs if an effect can be fully attributed to the public intervention to be evaluated. In the field of financial support programmes for female entrepreneurs, for example, it describes a situation where public money is fully additional to private money and not substituted by it.

Leverage effect

The concept of leverage is particularly important in relation to programmes promoting female entrepreneurship in the field of funding or finance. It is defined as the amount of private sector finance invested in an assisted project that has been induced by public spending.

The concepts of deadweight loss, additionality and leverage are illustrated in Box 10.

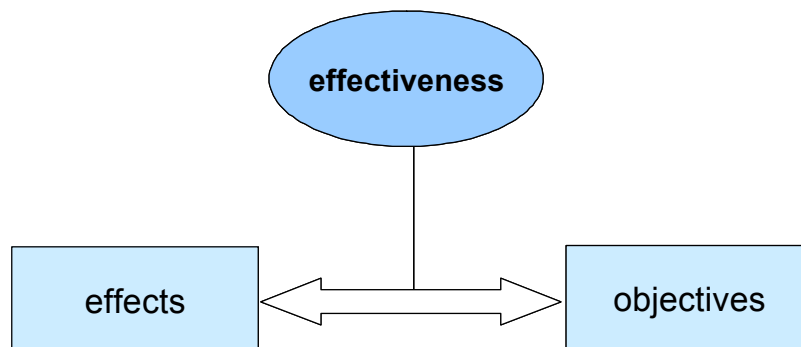


4.3.5 Judging the Effectiveness of a Programme

At a final stage of each evaluation, a judgement has to be made about the meaning of the evaluation findings in the context of the programme under investigation. As an evaluation is by definition a value judgement, the credibility of the evaluation is based entirely on the clarification of the references for this judgement, on the basis of which conclusions may be drawn as answers to the evaluation questions elaborated in the preparation phase of the evaluation and derived from the specific evaluation criteria (see section 3.1). A value judgement usually relates to one or more of these evaluation criteria, whereby the most predominating criteria applied in almost all impact evaluations are the *relevance* of a programme in relation to the needs of the addressees, the *coherence* of a programme referring to the degree of complementarity or synergy to be found within an intervention, the effectiveness of a programme in achieving its objectives, and the efficiency of an intervention, that is, its capacity to achieve the objectives at the lowest costs. The judgement of the former two criteria, the relevance and coherence of a programme, is dealt with in section 4.1. In this and the next section issues related to the judgement of the *effectiveness* and the *efficiency* of an intervention are described.

The judgement of a programme's effectiveness focuses on the extent to which the results and impacts of the programme contributed to achieving its specific and global objectives. Subsequently, effectiveness indicators relate the effects of a programme to its expected objectives (see also Box 11).

Box 11: Definition of the effectiveness of a programme



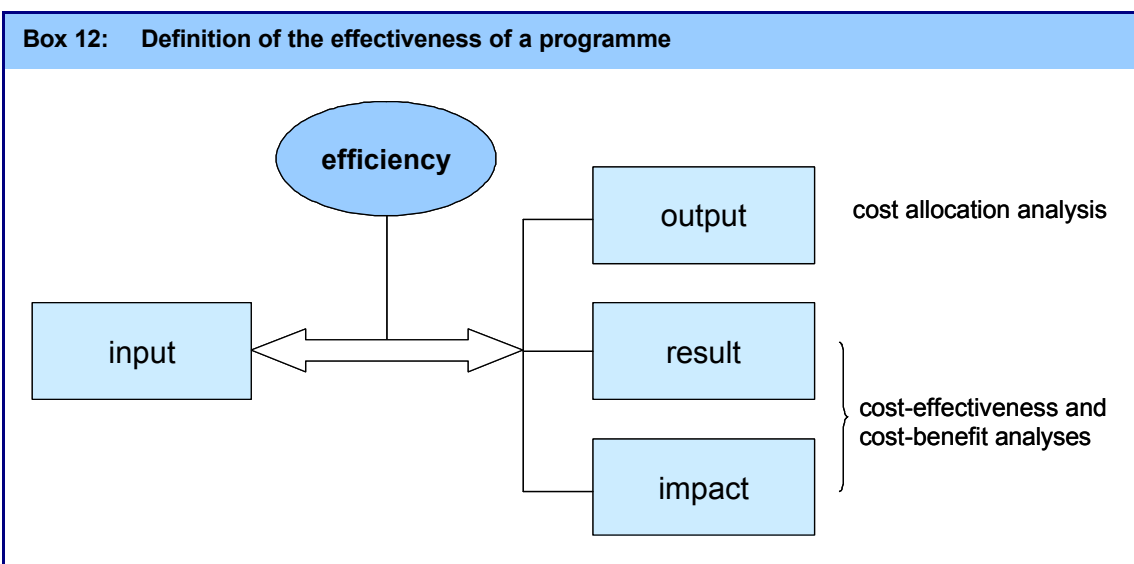
Assessing the effectiveness of a programme presupposes that the objectives have been unambiguously and realistically defined so as to make verification possible. In the case that objectives are unclear or highly general, the evaluation team will have to try to operationalise the objectives before judging the programme's effectiveness (see section 4.2).

Sample evaluation questions that relate to the effectiveness of a programme targeted at the promotion of female entrepreneurship are:

- ◆ *How well has the programme performed in terms of producing the expected outcomes?*
- ◆ *Have the objectives formulated been achieved? What are the successes and difficulties?*
- ◆ *Could more beneficial effects be produced by adopting other implementation measures in the scope of the programme?*
- ◆ *What is the influence of confounding factors outside the programme?*

4.3.6 Judging the Efficiency of a Programme

The judgement of a programme's efficiency relates to the productivity of the interventions, i.e. the relation between the output, result or impact obtained and the resources used (input) for a programme during a given period of time (see Box 12). Subsequently, efficiency indicators relate the effects of a programme to the resources mobilised for it.



Efficiency analysis may be undertaken at different levels of complexity. At the most basic level, an analysis of *cost allocation*, which concerns the relation between the inputs and the outputs of an intervention, shall be undertaken. This will usually be part of the monitoring process (see section 4.2.1) and allows programme managers to determine the true costs of providing a given unit of service, for instance. Cost allocation analysis also provides some of the basic information needed to conduct more ambitious cost analyses such as *cost-benefit* or *cost-effectiveness* analysis, for example, that aim at showing substantial longer-term social gains for participants or even cost savings for the public. Relating results, and even more so impacts, to the resources mobilised in the scope of an intervention may be rather complex and requires very sophisticated technical skills and training in methodology as well as in the principles of economics. However, efficiency analyses may essentially contribute to improving the understanding of a programme's operation, showing what levels of intervention are most cost-effective and revealing unexpected costs. In the following, the three techniques, cost allocation analysis, cost-benefit analysis and cost effectiveness analysis that may be applied for judging the efficiency of an intervention are briefly described.

Cost allocation analysis

For conducting a cost allocation analysis all direct and indirect costs of the programme to be evaluated have to be identified. Thereby, *direct* costs are those that benefit only the programme (e.g. salaries of programme staff or supplies and equipment used only for the programme) and *indirect* or "overhead costs" are those that benefit or are shared by more than one programme (e.g. several programmes run by an agency might share the same building and be served by the same bookkeeping and secretarial staff). Deciding on how to divide up the indirect (shared) cost pool among several programmes or measures can be complicated and technical. However, although cost allocation of indirect costs might be time-consuming, it is considered worth doing because of the increased information it delivers about the real costs of providing different types of services within a programme. Having determined the total costs of an intervention, i.e. the sum of the direct costs and the portion of indirect costs allocated to it, it is possible to calculate unit costs for each output of the programme.



Cost-effectiveness analysis

Cost-effectiveness analysis is characterised by its pragmatism, it focuses directly on the main effect of a programme in relation to its costs, purposely excluding other potential effects. Thus, cost-effectiveness analysis judges the effectiveness of an intervention in terms of a single criterion that is considered predominant or the key objective of the programme. On the basis of this criterion, it should be possible to compare the effectiveness of a programme with that of similar interventions in other regions or alternative actions focussing on the creation of similar results in the same region. Cost-effectiveness analysis is a comparative exercise; it assumes that a certain benefit or outcome is desired and that there are several alternative ways to achieve it. The basic question asked in cost-effectiveness analysis is:

Which of the alternatives is the cheapest or most efficient way to get the desired benefit?

Cost effectiveness analysis differs from cost-benefit analysis (see below) particularly in so far, as the effect of the intervention to be evaluated is expressed in physical quantities. While cost-benefit analysis always compares the monetary costs and benefits of a programme, cost-effectiveness studies often compare programmes on the basis of some other common scale for measuring outcomes (e.g. number of business start-ups by women). They address whether the unit cost is greater for one programme or action than another, which is often much easier to do, and even more informative, than assigning a monetary value to the outcome. Cost effectiveness analysis may constitute the first step in a benchmarking exercise, i.e. the qualitative comparison of a programme with another action implemented in a similar context and recognised as being excellent (a benchmark).

Cost-benefit analysis

The purpose of cost-benefit analysis is to determine whether the realisation of a programme is desirable from the point of view of society as a whole. The technique takes into account and assigns a monetary value to all the positive and negative effects that a programme has throughout its lifetime and compares all social and private costs of the intervention with its benefits, with a view to determining whether the benefits exceed the costs. The basic question asked in a cost-benefit analysis is:

Do the economic benefits of providing the programme outweigh the economic costs?

A key difficulty encountered in cost-benefit analysis is the valuation of social costs and benefits. Assigning monetary values to inputs and outcomes of programmes promoting female entrepreneurship or of other social interventions is rather complicated, and it is not always appropriate to do so.

Efficiency analyses can provide estimates of what a programme's costs and benefits are likely to be, also before its implementation. Ex-ante efficiency analyses may have to be based on very rough estimates of costs and expected benefits, however, if a programme is likely to be very expensive to implement, very difficult to "undo" once it is in place or very difficult to evaluate, even a rough estimate of efficiency may be quite valuable in the planning stage of the intervention.





5 ASSESSING THE QUALITY OF AN EVALUATION

Although the quality of an evaluation is often judged on the sole basis of the final report, it largely depends on the way in which the entire evaluation process was managed, from the commissioning of the evaluation to the conduction of the work, the drawing of conclusions, the formulation of recommendations and the communication of results. It is necessary to assess the quality of an evaluation in order to verify whether the evaluation team satisfied the commissioners' requirements and to ensure that the evaluation can stand up to criticism that may arise when judgements on success and failure are made. In particular, it has to be decided whether the collection and analysis of data is sufficient or must be improved, whether the conclusions drawn from the evaluation are acceptable in view of the proof gathered, and whether the report answers all the questions mentioned in the terms of reference or whether further work needs to be undertaken. A grid for assessing the quality of an evaluation is provided in Box 13.

Box 13: Grid for a synthetic assessment of the quality of evaluation work

With regard to this criterion, the evaluation report is:

inacceptable
acceptable
good
excellent

1. Meeting needs: Does the evaluation adequately address the requests for information formulated by the commissioners and does it correspond to the terms of reference?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Relevant scope: Have the rationale of the programme, its outputs, results, impacts, interactions with other policies and unexpected effects been carefully studied?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Defensible design: Is the design of the evaluation appropriate and adequate for obtaining the results (with their limits of validity) needed to answer the main evaluative questions?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Reliable data: Are the primary and secondary data collected or selected suitable? Are they sufficiently reliable compared to the expected use?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Sound analysis: Are the quantitative and qualitative data analysed in accordance with established rules and are they complete and appropriate for answering the evaluative questions correctly?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Credible results: Are the results logical and justified by the analysis of data and by interpretations based on carefully presented explanatory hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Impartial conclusions: Are the conclusions just and non-biased by personal or partisan considerations, and are they detailed enough to be implemented concretely?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Clear report: Does the report describe the context and goals, as well as the organisation and results of the evaluated programme in such a way that the information provided is easily understood?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In view of the contextual constraints bearing on the evaluation, the evaluation report is considered to be	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



The quality assessment grid presented in Box 13 can be applied to ex ante, intermediate as well as ex post evaluation. It has been developed within the framework of the MEANS programme (see References at the end of this guide), with the help of experts from several European countries. The grid is based on eight quality criteria that have already been used by the European Commission as well as by national authorities for assessing the quality of a number of evaluation reports. The global assessment of an evaluation is established synthetically after the final report has been analysed in terms of these eight criteria which are briefly described below:

1. Meeting needs

It is necessary to check whether the evaluation has answered the questions in the terms of reference - which ideally group together the diverse requirements of all stakeholders of the evaluation - satisfactorily and whether the report covers additional requests for information, as well as new questions which might be essential for the commissioners. When assessing this criterion, the following questions should receive particular attention:

- *Has the evolution of global and specific objectives of the programme been analysed?*
- *Does the report cover the entire programme? If not, is the selection justified in regard to the priorities stated by the commissioners?*
- *Does the evaluation provide useful feedback for the managers of the programme?*
- *Does it include lessons on successes and failures which may be of interest to other regions or countries?*
- *Have all relevant impacts been observed?*

2. Relevant scope

In order to check the relevance of the scope of the evaluation, it is necessary to assess the following questions, whereby the relative weight of each question depends on the timing of the evaluation, i.e. whether the evaluation was conducted ex ante, intermediate or ex post:

- *Does the report evaluate in sufficient detail the target group's expectations and the corresponding relevance of the programme's objectives and strategies as well as the internal and external coherence of the programme (ex ante evaluation)?*
- *Have the essential characteristics of the programme been well described and have the problems and successes in the implementation of the programme been properly clarified (intermediate evaluation)?*
- *Have the results and impacts of the programme necessary to judge the extent to which its objectives are achieved been included in the evaluation and hasn't the evaluation overlooked other potential or future results or impacts, as well as any unexpected, but significant effects which may exist (ex post evaluation)?*

3. Defensible design

This criterion is related to the technical qualities of the evaluation. Three types of questions have to be asked:

- *Have the relevant knowledge and existing literature been collected and used wisely?*
- *Are the construction of the evaluation design and the choices of the data collection techniques justified for answering the evaluation questions properly?*
- *Were the reference situations (e.g. counterfactual situation) chosen properly for making valid comparisons?*



Any evaluation report must include a description of the evaluation method used and must clearly define the sources of data applied. Similarly, the limits of the method and techniques used must be thoroughly explained. Thus, it is also necessary to check whether:

- the method is described in enough detail to judge the technical quality of the evaluation
- the validity of the data collected and the techniques used is clearly indicated
- the available data correspond to the techniques used
- the evaluation adequately analyses relations of cause and effect for the most essential evaluation questions

4. Reliable data

Evaluators use secondary data from the monitoring system and from other sources of information as well as primary data which they have collected for the evaluation. In order to assess the reliability of the data used, it is, thus, necessary to examine the following questions:

- *Have available sources of information been identified and has the reliability of the data been checked?*
- *Has the information taken from the monitoring system and previous studies been used optimally?*
- *Were the techniques used to collect the data complete and suitable for answering the evaluation questions and have the survey samples or case studies been selected in relation to established criteria?*
- *Have the main data collection techniques been implemented with appropriate tools and in such a way as to guarantee an adequate degree of reliability and validity of results?*

5. Sound analysis

In order to check the soundness of the analysis, it is necessary to assess whether the methods of qualitative and quantitative analysis used are relevant as regards the type of data collected, and whether the analysis has been carried out correctly, i.e. according to the instructions in the relevant technical manuals.

6. Credible results

The credibility of results is defined here as the fact that they follow logically from and are justified by the analysis of data and interpretations based on carefully presented explanatory hypotheses. The balance between the internal validity (i.e. the absence of technical bias in the collection and processing of data) and the external validity (i.e. the representativeness of results) must be justifiable. When the evaluation is based on the conduction of case studies, for example, it has to be checked, whether the interpretative hypotheses and extrapolations are reasonable and whether the selection of cases and samples makes it possible to obtain relevant findings. Furthermore, the following questions need to be asked:

- *Are the results of the evaluation credible in the sense of reflecting an acceptable compromise between the reality of the programme, as perceived by the actors, and the reality described by the observed or estimated figures and facts?*
- *To what extent were the stakeholders of the programme involved in the process of interpreting empirical facts?*
- *Were the results presented impartially to the stakeholders?*



7. Impartial conclusions

In order to judge whether the conclusions and recommendations drawn on the basis of a value judgement of the evaluation results are fair, free of personal considerations and detailed enough to be implemented concretely, it is necessary to check whether:

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

8. Clear report

The evaluation report is the end product of the evaluation and one of the main means of diffusion and communication of the evaluation results. It is, thus, necessary to check whether:

- the report is written clearly for its addressees and whether it is set out logically
- the 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

Of course, there is no universally applicable structure for an evaluation report. However, it is important that the structure of the report meets the needs of the commissioners as well as of the principle stakeholders of the evaluation and that it contains a clear and concise summary of the evaluation results, presenting the main conclusions and recommendations in a balanced and impartial matter and easy to read without the need to refer to the rest of the report. Box 14 presents an example of a typical structure for an evaluation report.

The final assessment of the quality of the evaluation must take into account the constraints of the evaluation. Two types of constraints can be distinguished depending on whether the commissioning of the evaluation or the management of the evaluation is concerned. Constraints relating to the commissioning procedure are: the clarity and precision in the terms of reference, the volume of budgetary inputs, the time allocated to the evaluation, and the involvement of the commissioners during the implementation of the evaluation. Constraints relating to the programme to be evaluated and its management may result from a lack of clarity of objectives, inadequate quality of the data generated by the monitoring system or crises or ruptures which may occur during programme implementation and which may have an impact on the evaluation. It is up to the commissioners to limit the effects of the constraints related to the commissioning procedure and up to the evaluation team to anticipate technical and operational problems before embarking on the evaluation process.



Box 14: Example of an evaluation report structure**Executive summary**

- overview of the entire report in no more than five pages
- discussion of the strengths and weaknesses of the evaluation method

Introduction

- short introduction to the evaluation report, including a brief outline of the programme at stake and the evaluation design
- possible constraints in what the evaluation was able to do
- description of the structure of the report

Description of the programme and main evaluation question

- description of the programme at stake in terms of needs addressed, objectives, activities and components, delivery mechanisms, and resources used for its implementation.
- description of the context in which the programme operates
- purpose of the evaluation in terms of scope and main evaluation questions

Research methodology applied

- description of the evaluation method
- description of the types of data collected and techniques used for collecting the data, including, for example, how samples were selected and how representative they are
- short explanation of the techniques used to analyse the data

Evaluation results

- results of the evaluation

Conclusions

- clear and precise description of the conclusions based on the overall assessment and answering the evaluation questions asked in the terms of reference
- possible policy recommendations derived from the evaluation





6 DISSEMINATING AND USING EVALUATION RESULTS

If an evaluation shall add real value in the institutional and decision-making spheres, its conclusions must be disseminated correctly and must be viewed favourable by the potential users of the evaluation at stake. These might range from key policy makers and decision makers, programme managers and commissioners of evaluations to programme beneficiaries, the academic community with a general scientific interest in the evaluation or other interest groups, such as organisations or individuals in the immediate environment of the programme to be evaluated. By carefully planning the dissemination strategy, which encompasses the whole range of activities by which the information contained in evaluation reports is made available to the wider audience, in an early stage when the terms of reference are drawn up, the use of evaluation can be optimised. Thereby, the distribution of the evaluation report itself is not the only source of communicating evaluation results; it may be supplemented by other forms of communication, such as conference presentations, workshops and seminars, press releases or the distribution of the executive summary, for example. The establishment of a clear communication plan considering the diverse information needs of the target audience and synchronised to the time planning of the evaluation may facilitate the dissemination process and enhance the utilisation of the evaluation results.

The specific use of the evaluation findings will, of course, depend on the overall purpose of the evaluation and the questions addressed in it. However, three general forms of using evaluations may be distinguished:

- *Improving the programme*

The evaluation may be used to improve the programme, whereby evaluation findings should support decisions and actions to best do so. Specific findings might be used to identify strengths and weaknesses of the programme or provide strategies to improve the programme midway or at the time of its renewal.

- *Generating new knowledge*

Any evaluation will offer its users an opportunity to discover, to learn and possibly to understand how and why a programme works, for whom, and in what circumstances. The evaluation may provide information about general principles of good practice, connections between underlying theories and practice, and sometimes lead to new and enhanced theories about human and organisational development. These types of findings can be used to collaborate, share and learn across programmes with common themes and principles.

- *Judging success and failure*

Evaluation may be used to draw conclusions on the success and failure of a particular programme or parts of it and to decide on a programme's future, determining the likelihood of continued funding or decision making about the programme's expansions. This form of use is essential when evaluation serves as accounting to political authorities and the public at large.





REFERENCES

Baker, J. (2000): *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Directions in Development. Washington, D.C., World Bank.

Bellini, N. (2002): *Linking Policy Evaluation and Learning: Towards a Re-appraisal of the Role of Business Support Services in the European Regional Policies*. In: Regional Studies Association International Conference on Evaluation and EU Regional Policy: New Questions and New Challenges. Aix en Provence.

Canadian International Development Agency, CIDA (1997): *Guide to Gender-Sensitive Indicators*. Hull, Quebec.

Canadian International Development Agency, CIDA (1997): *The Why and How of Gender-Sensitive Indicators: A Project Level Handbook*. Hull, Quebec.

Council of Europe (1998): *Gender Mainstreaming: Conceptual Framework, Methodology and Presentation of Good Practices*. Final Report on Activities of the Group of Specialists on Mainstreaming. Strasbourg, EG-S-MS (98) 2.

Curnan, S. / LaCava, L. / Langenburg, D. / Lelle, M. / Reece, M. (2000): *Kellogg Foundation Evaluation Handbook*. Michigan, Evaluation Unit of the W.K. Kellogg Foundation.

European Commission, DG Enterprise (2002): *Support Services for Micro, Small and Sole Proprietor's Businesses*. Final Report. Vienna, Austrian Institute for Small Business Research (IfGH).

European Commission, DG Enterprise (2001): *Creating Top-Class Business Support Services*. Commission Staff Working Paper. Brussels, SEC(2001)1937.

European Commission, DG Budget (2001): *Ex Ante Evaluation – A Practical Guide for Preparing Proposals for Expenditure Programmes*. Brussels.

European Commission, DG Agriculture (2001): *Common Market Organisations in Agriculture: Current Evaluation Practice Guide*. Brussels.

European Commission; DG Enterprise (2000): *Young Entrepreneurs, Women Entrepreneurs, Ethnic Minority Entrepreneurs and Co-entrepreneurs in the European Union and Central and Eastern Europe*. Summary Report. Middlesex, Centre for Enterprise and Economic Development Research (CEEDR).

European Commission, DG Regional Policy (2000): *The New Programming Period 2000-2006: Methodological Working Papers*. Working Paper 3. Indicators for Monitoring and Evaluation: An Indicative Methodology. Brussels.

European Commission, DG Regional Policy (2000): *The New Programming Period 2000-2006: Technical Papers by Theme*. Technical Paper 3. Mainstreaming Equal Opportunities for Women and Men in Structural Fund Programmes and Projects. Brussels.

European Commission, DG Regional Policy (1999): *Evaluating Socio-Economic Programmes*. MEANS Collection Vol. 1 to 6.

European Commission, DG Budget (1997): *Evaluating EU Expenditure Programme: A Guide. Ex Post and Intermediate Evaluation*. 1st Edition. Brussels.



European Commission, DG Employment and Social Affairs (1996): *Communications: Incorporating Equal Opportunities for Women and Men into all Community Policies and Activities*. Brussels, COM(96)67final.

European Commission, DG Regional Policy (1995): *Common Guide for Monitoring an Interim Evaluation of CSFs, SPDs and Operations*. Structural Funds. Brussels.

Government of Western Australia, Public Sector Review Division (1995): *Managing the External Evaluation Process – A Practical Guide*. Perth.

Kantor, P. (2000): *Promoting Women's Entrepreneurship Development based on Good Practice Programmes: Some Experiences from the North to the South*. In: Series on Women's Entrepreneurship Development and Gender in Enterprises (WEDGE). Wisconsin Madison.

Ministry of Foreign Affairs Danida (1999): *Evaluation Guidelines*. 2nd Edition. Copenhagen.

Muraskin, L.D. (1993): *Understanding Evaluation: The Way to Better Prevention Programs*. Westat Inc. Newton.

O'Reilly, J. and Schönberg, K. (1996): *International Handbook of Labour Market Policy and Evaluation*. Cambridge, University Press.

Organisation for Economic Co-operation and Development, OECD (2000): *Evaluation of SME Policies and Programmes*. In: OECD Small and Medium Enterprise Outlook. Enterprise, Industry and Services. Paris, OECD.

Organisation for Economic Co-operation and Development, OECD (2000): *Improving Evaluation Practices. Best Practice Guidelines for Evaluation and Background Paper*. Paris, OECD, PUMA/PAC(99)1.

Patton, M.Q. (1995). *How to Use Qualitative Methods in Evaluation*. Newbury Park, Sage.

Pawson, R. / Tilley, N. (1997): *Realistic Evaluation*. Thousand Oaks, Sage.

Pimminger, I. (2001): *Handbuch Gender Mainstreaming in der Regionalentwicklung. Einführung in die Programmplanung*. Wien.

Prennushi, G. / Rubio, G./ Subbarao, K. (2000): *Monitoring and Evaluation*. In: A Sourcebook for Poverty Reduction Strategies. Washington D.C., World Bank.

Rossi, P.H. / Freeman, H.E. (1993): *Evaluation. A Systematic Approach*. 5th Edition. Newbury Park, Sage.

Rossi, P.H. / Freeman, H.E. / Hofmann, G. (1988): *Programm Evaluation: Einführung in die Methoden angewandter Sozialforschung*. Stuttgart, Enke.

The Urban Institute (1997): *Evaluation Guidebook for Projects Funded by S.T.O.P. Formula Grants under the Violence against Women Act*. Washington D.C.

United Nations Development Programme, Evaluation Office (2002): *Handbook on Monitoring and Evaluating for Results*. Yellow Handbook. New York.

U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention by Community Research Associates, Inc. (1989): *Basic Outcome Evaluation and Comparative Outcome Evaluation*. In: Office of Juvenile Justice and Delinquency Prevention. Evaluating Juvenile Justice Programs: A Design Monograph for State Planners. Washington, D.C.



NOTES

