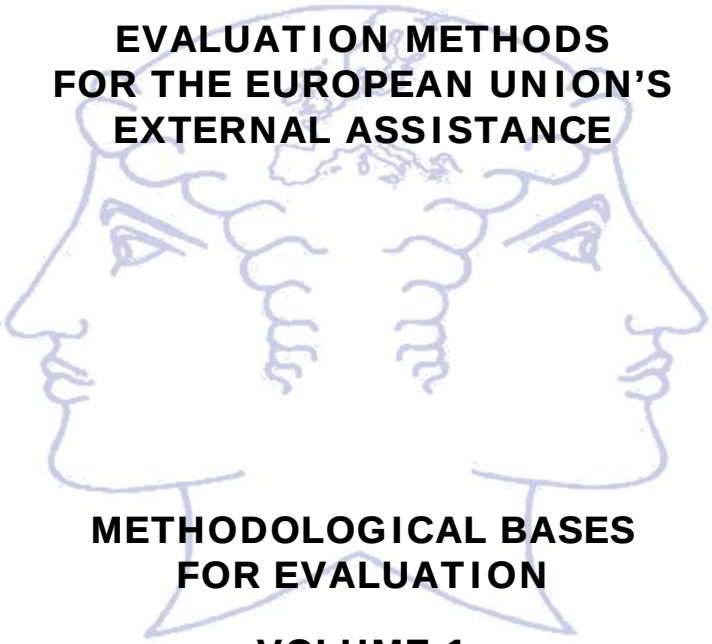
Directorate General External Relations

Directorate General Development

EuropeAid Co-operation Office

Joint Evaluation Unit

EVALUATION METHODS FOR THE EUROPEAN UNION'S EXTERNAL ASSISTANCE



METHODOLOGICAL BASES FOR EVALUATION

VOLUME 1

A great deal of additional information on the European Union's external assistance can be accessed through the Europa server: *http://ec.europa.eu/europeaid/index_en.htm*

Luxemburg: Office for Official Publications of the European Communities.

ISBN: 92-79-00681-9

Printed in France, 2006.

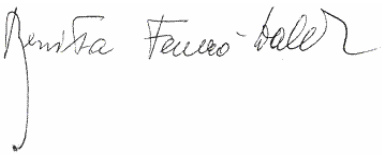
Foreword

The constant improvement of evaluation practices is essential for the European Commission, particularly in the areas of external assistance and development. In addition to accounting to European taxpayers for the sound use of public funds, there is a need to learn lessons from past interventions. For these purposes, the European Commission has to be able to rely upon credible evaluation methods that ensure the validity of conclusions and recommendations.

The reform of the management of external aid, launched in 2000, included a reassertion of the importance of evaluating the results of regional and sectoral policies and programmes, as well as the effectiveness of programming. Evaluation is now a key factor of success of EU external aid policies as a whole, and is incorporated into the programming cycle.

Between 2002 and 2005, the Joint Evaluation Unit for the Directorates-General EuropeAid, Development and Foreign Affairs of the European Commission developed an evaluation approach aimed at meeting the needs defined in 2000. The results of this methodological effort are available on the Commission's website and are summarised in four booklets. These documents constitute a useful means to better understanding and mastering the evaluation approach which has been implemented within the EuropeAid Cooperation Office. I am pleased to note that this approach is already a benchmark among donors.

The European Commission's approach is publicly available and will be of particular interest to all those involved in official development assistance and co-operation with third countries.



Benita Ferrero Waldner

Synthesis of comments by Expert Panel on the evaluation manual of EuropeAid

Background

These notes were made after the draft manual for evaluation was presented to us in the expert panel in early September 2005. The comments refer to the material sent to the panel members in August 2005. The comments refer only to the draft manual, and they do not take into account other formulations of evaluation policy or presentation of the practices of European development cooperation. The panel met on the 16th of September, and before that each of us had spent some days reviewing the draft manual. Our meeting on the 16th lasted 2 hours, and following that we have had e-mail exchanges over these notes.

Overall impression of the manual

The draft manual is a very comprehensive set of documents and it sets out the evaluation process in great detail. It reflects a committed effort to clearly present and structure the evaluation process and to produce guidance that responds to the needs of people in the organisation.

The manual could also be very useful for other stakeholders who in various ways are affected by the evaluation processes of EuropeAid, for example consultants or representatives from authorities in partner countries.

The manual is well balanced. The major components parts are of equal weight, and no aspect of the evaluation process gets more attention than its due part of the process merits. Most of the sections that constitute the manual are of a similar length, of around 1 and 2 pages. The manual often treats complex concepts and processes with admirable brevity. At the bottom end of the hierarchical structures the texts are sometimes longer, even up to a dozen pages (Sections on evaluation tools can go up to 40 pages). However, this is to be expected as the longer texts are at the end of a line of quest and this poses no problem.

The manual addresses different readers and it can be approached differently depending on who you are and what your interests are.

This is useful and it makes the manual as a whole more flexible and adaptive.

Most of the content of the manual reflects an emerging consensus in the evaluation community on how evaluation should be undertaken. It is up-to-date in 2005 and should be able to serve EuropeAid for many years to come.

Our comments should be seen in the light of an overall impressive product which reflects a comprehensive grasp of evaluation as a phenomenon and substantive skills in encoding this knowledge and communicating it to others through the manual and in designing a web-based source of organisational learning.

Issues of policy

We have some comments that relate to the overall evaluation policy as it is reflected in what is written – or not written – in the manual. Whether the manual ought to be amended to take account of our observations depends on the position that EuropeAid takes on the substance of these issues.

Practical guidance and advice on joint evaluation

It is often recommended that funding agencies should join forces in evaluation. There are many reasons for that, not least that the evaluation process becomes less of a burden for developing countries, whose project personnel and other officials otherwise have to spend too much time with external evaluation missions. Another reason is that projects and programmes are frequently financed and/or delivered jointly, and therefore it makes sense to evaluate them jointly. It has been strongly recommended by the OECD/DAC Working Group on Evaluation that coordination between member states should be increased.

Development cooperation is changing and though budget support, sector assistance and programme approaches have supplemented traditional project assistance for the past two to three decades, the so called newer forms are continuously becoming more prevalent. It is also the broader approaches to cooperation that are most suitable for cooperation between funding agencies and where this also would have the highest effect in partner countries.

However, cooperation between funding agencies in evaluation would require some compromises. As all the agencies concerned have their own ways of codifying and regulating the evaluation process, joint evaluation would often need a significant element of

mutual adjustment. The manual does not contain any advice or guidance on how evaluation managers should approach joint evaluation. In the future, it is likely that evaluation managers will increasingly find that they manage evaluations together with colleagues from other agencies, and hence they must adapt. This is a difficult task, and the question is whether there are some aspects of the process that are less “negotiable” than others, and if so how they could collaborate with others.

In our opinion, the evaluation manual could benefit from an additional section on joint evaluation, and furthermore we would suggest that this is described not as a problem and a further complication, but as an opportunity to strengthen the evaluation process, achieve economies of scale, and that will lead to more effective collaboration for the host country.

Approach to participatory evaluation

Yet another significant development in the field of evaluation research lies in participatory evaluation. The concept arose through research on the utility and use of evaluation findings. In cases where the primary function of evaluation is formative, it has been found particularly useful to involve various stakeholders/participants in the process of defining questions, developing methods, collecting data and drawing conclusions. When there is a stronger sense of ownership of the intellectual outcome of the process, it is more likely that stakeholders will act on the recommendations. Stakeholders in this sense refer organisations that are engaged in development cooperation; such as NGOs, consulting firms, public authorities at central and decentralised levels, etc.

When other actors are engaged in the process, it would again often be necessary to adjust the process. The reference group that is defined in the manual may, for example, have other roles and functions, and there might be supplementary bodies. The evaluation team may also have other roles and may need other competencies, perhaps more in the nature of facilitators to the process rather than as operators. As in respect of joint evaluation, we think that the manual may contain words of advice on participatory evaluation strategies, encouraging evaluation managers to experiment with participatory approaches and advising them how to adjust and adapt the process to also accommodate other actors.

Ethical standards

The evaluation community has devoted much attention to the question of what constitutes evaluation quality. The Programme Evaluation Standards of 1994 (The Joint Committee on Standards for Educational Evaluation. The Program Evaluation Standards. Thousand Oaks, CA Sage Publications, Inc. 1994) is one of the path-breaking and most influential publications on the subject. It discusses quality in four dimensions; utility, feasibility, accuracy and propriety. The manual that we have reviewed deals thoroughly with aspects of the evaluation process that contribute to feasible, accurate and useful evaluation, but it does not contain much material on evaluation ethics.

There are good reasons to believe that ethical subjects will – in the long run – prove particularly difficult in evaluation of development cooperation. The asymmetrical power relations, the prevalence of donor- recipient modalities of thinking and acting, the often perverse incentive systems around aid, and the cross-cultural differences contribute to make aid evaluation difficult and subject to intricate ethical choices. The manual could be supplemented with a section on what ethical standards mean and why they are so important in aid. It could also outline some of the ethical pitfalls that evaluation managers and evaluation teams are likely to face – and fall into.

Integration with related systems

As far as we have understood it, EuropeAid as an organisation applies a system of Results Based Management (RBM). It is not quite clear how this system as a whole, and the decision-making in this system, will relate to the evaluation system. There is ambiguity as to whether these materials are created to work within or outside of a focus on and commitment to the perspective of results-based management. There are places that hint of an acceptance of RBM as the conceptual Boxwork for this material and then there are multiple places where the content suggests it is not—especially in the sections on indicators and targets. This needs attention and an explicit statement in the beginning of the materials as to the relation of what is here to the RBM perspective.

We would also think there is a need for further discussion on how the materials presented here link in explicit ways to monitoring and also to auditing. There is no discussion of the linkages that

would be necessary to create a monitoring and evaluation (M&E) system in a country, for programs, or even for projects. In fact, the materials go out of their way to make distinctions between evaluation and both monitoring and auditing, but nothing is made of the similarities and linkages that could come from greater coordination among these approaches. In the past, EU evaluations have often been constrained by the limited usefulness of monitoring data. It would be a pity if this opportunity were not taken better to co-ordinate these streams of activity.

Concluding remarks

It has been useful for us to go back and attempt an overview of the website materials after more than 2 years of effort. At the core of the task set for the expert group lies an assessment of the 'quality' of these materials. 'Quality', however, is a concept notoriously open to both broad and narrow interpretations. There could be:

- a narrow definition of scientific quality, and
- a broad definition of quality as the ultimate usefulness of the material to the range of its intended users

With respect to 1 (scientific quality/technical adequacy) EuropeAid has done well. There are some criticisms to be made but, in general, both the comprehensiveness of these texts and their technical accuracy/awareness are sound.

With respect to 2 (usefulness to stakeholders), however, we suggest that there is still some way to go. 'Motivation' is almost as important as 'credibility' or 'dissemination'. But we have seen little so far in this website to motivate evaluation managers. The material equips them to do a job, but does little to convince them that the job is valuable, creative, interesting and generally worth doing. As we understand it, many of the evaluation managers in the Delegations will have plenty of other responsibilities, so their evaluation business is in constant danger of becoming just an additional chore – to be minimized. Whilst no website can change that by itself, it can at least show some sympathy for the problem.

Experts

Kim Forss, Andante, Stockholm

Kim Forss works for Andante - tools for thinking AB. His work is dedicated to research and consultancy in the field of evaluation research, policy research on international relations and organisational development. He is the author of communications dealing with quality and use of evaluations results in a perspective of improved democratic management.

John Mayne, Ottawa

John Mayne is an independent advisor on public sector performance. Previously, he was with the Office of the Auditor General of Canada and the Treasury Board Secretariat of Canada. He has worked extensively on evaluation quality assessment. He has also edited several books in the areas of program evaluation, including an international book on evaluation and performance measurement.

Christopher Pollitt, Rotterdam University

Christopher Pollitt has a long record in research and teaching in evaluation and public management, first in Brunel University (UK) and currently in Rotterdam University (NL). He has been part of many OECD research networks in these areas. He has been the chairman of the European Evaluation Society. He has produced a critical analysis of guidelines.

Ray Rist, World Bank – OED, Washington

Ray Rist currently works as a methodological adviser for OED (World Bank). He has worked previously in the US General Accounting Office. He has an extensive knowledge of evaluation practices worldwide.

Pierre Spitz, INRA, Paris

Pierre Spitz works for the Institut National de la Recherche Agronomique (INRA) in Paris. His area of expertise concerns rural development. He used to work for UN agencies for the last decade.

TABLE OF CONTENTS

I. INTRODUCTION

Overview

The European Commission has developed and formalised a methodology for evaluating its external assistance, in which the priority is on results and impacts. The aim is thus to maintain the quality of its evaluations on a par with internationally recognised best practice.

In the past, the evaluation of European external assistance focused on projects and on certain programmes. The current methodological guidelines are designed to facilitate the move towards an evaluation practice focused more on programmes and strategies. It is intended mainly for:

- evaluation managers at European Commission headquarters and in the Delegations,
- external evaluation teams.

The methodology is also made available to all European external aid partners, as well as the professional evaluation community.

It is available in three languages (English, Spanish and French) and in two forms, optimised for reading and for navigation on the Internet, respectively.

The Internet version includes numerous examples and in-depth analyses. It is available on the European Commission website:

http://ec.europa.eu/europeaid/evaluation/index.htm

The printed version consists of four volumes. This first volume presents the basic concepts and their articulation. The second volume is a handbook for "Geographic and Thematic Evaluation". It pertains to the evaluation of the entire set of Community actions on the scale of a country or region, and the evaluation of all actions relative to a sector or a theme on a global scale. The third volume is a handbook for "Project and Programme Evaluation". It concerns large projects, pilot projects, multi-country programmes and any other project or programme for which an evaluation is required. The fourth volume "Evaluation Tools" presents the main techniques available for structuring an evaluation, collecting and analysing data, and assisting in the formulation of value judgements.

Methodological bases

This volume presents most of what needs to be known in order to manage or to carry out an evaluation, using a standardised terminology in three languages (English, Spanish and French) that contributes towards the OECD Development Aid Committee's efforts at harmonisation (see p. 92). It offers complete, concrete and tested solutions to many common problems, including the most difficult ones, such as clarifying the intervention rationale, formulating evaluation questions, designing a method and analysing the impacts.

The volume consists of two main parts:

- The process: object of the evaluation; timing; utilisation; players of the evaluation and their roles.
- On methods: intervention strategy; evaluation questions (usefulness, feasibility, formulation); judgement references (criteria and indicators); methodological design; data collection and analysis; value judgement (conclusions, lessons and recommendations); and, finally, quality assurance.

II. THE EVALUATION PROCESS

What is evaluation?

OECD - DAC

- The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision–making process of both recipients and donors.
- Evaluation also refers to the process of determining the worth or significance of an activity, policy or program. An assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention.

Dg Budget

- Judgement of interventions according to their results, impacts and needs they aim to satisfy (The Communication on Evaluation: SEC(2000) 1051)

The main purposes are as follows:

- To contribute to the design of interventions, including providing input for setting political priorities
- To assist in an efficient allocation of resources
- To improve the quality of the intervention
- To report on the achievements of the intervention (i.e. accountability).

The following chapters are dedicated to the evaluation process, successively focusing on: the subject of the evaluation (what is evaluated), the timing, the utilisation, the players of the evaluation and their roles.

1 Subject of the evaluation

Summary

Project, programme, strategy

The assessed intervention may be simple, for instance, it may be a project aimed at generating a specific and direct effect on a given public, or, on the contrary, it may be complex, such as a strategy generating its effects by means of other lower level interventions. Depending on the intervention's degree of complexity, the questions addressed and the evaluation method may vary.

Sectors, themes and cross-cutting issues

Public interventions are usually classified by sector or theme. A particular intervention or set of interventions can be evaluated from the point of view of a cross-cutting issue.

Scope of the evaluation

The scope is everything that is subject to the evaluation investigations. Its perimeter is defined in terms of several dimensions: geographic, temporal and institutional.

A second and larger perimeter specifies related actions and elements of context to take into account.

1.1 Project, programme, strategy

The term « intervention » refers to what is evaluated, that is to say one or several projects, programmes, strategies or any other kind of external assistance action.

Evaluation of a project

A project is an indivisible operation, delimited in terms of time schedule and budget, and usually placed under the responsibility of a single operator. Homogeneous projects are the easiest to evaluate (see Box 1). They are characterised by:

- One main activity, implemented in a single context, intended for a single target group
- A single expected result directly accruing for that group
- A single expected impact at a more global level.

The evaluation of a simple project is facilitated in that it focuses on a single cause (the main activity) and a single effect (expected result for the target group).

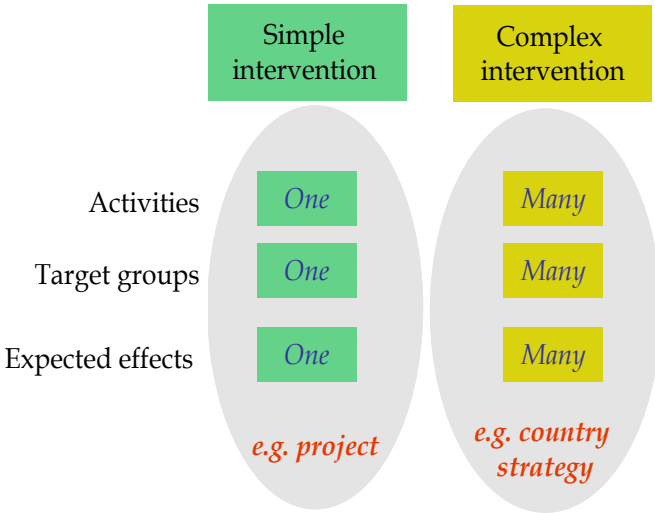
A project evaluation often focuses on issues of interest to managers, operators and the targeted group.

Evaluation of a programme

A programme is a set of simple, homogeneous interventions clustered in order to achieve global objectives. It is delimited in terms of time schedule and budget and usually placed under the responsibility of a monitoring committee.

A homogeneous programme is made up of parallel and similar projects. The evaluation of a homogeneous programme can consist in a synthesis of project evaluations or an extrapolation from a sample of projects.

Box 1 –Simple versus complex interventions



Evaluation of complex interventions

Complex interventions (such as a country strategy) gather diverse but complementary interventions and serve a common objective. They are designed with:

- Several activities (subsidies, directs investments, technical assistance, etc.), implemented in several contexts (several geographical areas; various sectors …) targeting various groups
- With many expected results for these target groups
- And several expected impacts at a more global level.

The evaluation of a complex intervention is neither the sum nor the synthesis of the evaluations of its components. It focuses on questions that are relevant for policy makers (synergy effects, relevance of the allocation of resources between the components, contribution to the achievement of the global objectives).

1.2 Sectors, themes and cross-cutting issues

Public interventions are usually classified by sector or theme. A particular intervention or a set of interventions can be evaluated from the point of view of a cross-cutting issue, which is sometimes referred to as a thematic evaluation.

Sector

Interventions are classified in a particular sector according to the nature of the activities and outputs. For instance:

- A support to training primary education teachers is classified in the education sector
- Advisory services provided to farmer groups are classified in the agriculture sector

Cross-cutting issues

A cross-cutting (or horizontal) issue is defined in relation to the impacts and not to the outputs, as shown by the two examples in the Box 2:

Box 2 – Difference between sectors and cross-cutting issues

	Example 1	Example 2
Output	Teachers trained	Advice provided to groups of farmers
Sector	*Education*	*Agriculture*
Impact	Girls benefit from a better access to education	New breeding practices that prevent desertification
Cross-cutting issue	Gender	Environment

Theme

The term "theme" is sometimes used in the sense of a sector and sometimes in the sense of a cross-cutting issue. It is therefore advisable to redefine it every time it is used.

Warning!

The same term can denote a sector or a cross-cutting issue. For instance, human rights can be:

A sector, when an evaluation focuses on a training project for staff of the law courts

A cross-cutting issue, when an evaluation focuses on improving access to the Internet in poor rural areas and the impact in terms of reducing human rights violations.

1.3 Scope of the evaluation

What is submitted to a value judgement is the scope of the evaluation. The scope is delineated in various dimensions such as the territory concerned, the period under consideration and the regulatory Boxwork, for example:

- All the funds allocated by the European Union to Latin America since 1998
- Local development actions initiated in Rwanda over the past four years.

A second perimeter specifies the related actions and elements of context to be taken into account, especially for the analysis of external coherence/complementarity. This is the extended scope, as opposed to the central scope.

Delineation of the extended scope

Related policies

- The interventions of national authorities or other donors covering the same territories
- … or targeting the same groups
- … or addressing the same problems

Co-funding

- interventions co-funded with national authorities
- interventions co-funded with other donors

Other relevant elements of the context.

The central scope is specified in the terms of reference and the extended scope in the inception report.

2 Timing of the evaluation

SUMMARY

Case of a stand-alone intervention

An evaluation can be performed before, during or after the evaluated intervention. Depending on the timing, its purpose and use will differ.

Ex ante evaluation is an input into the formulation of the intervention.

An evaluation during or at the end of the intervention cycle is intended to draw lessons from the first years of the implementation and to adjust the contents of the ongoing intervention in relation to realities in the field and/or contextual developments.

The ex post evaluation is mainly concerned with assessing achieved impacts, identifying and judging unexpected impacts and assessing the sustainability of the intervention's benefits.

Case of a series of interventions

Many interventions are characterised by successive cycles and show a relative continuity between cycles.

In light of this, several cycles need to be considered when choosing the timing of the evaluation.

It is important to take into account the timing of the evaluation in order:

- To optimise the resources allocated to the evaluation by launching it at the time it is likely to have the most added value
- To meet the needs of the main users of the evaluation at the most appropriate time
- To ensure that a critical mass of results and impacts are already materialised in the field and they lend themselves to data collection
- To avoid conflicts with any concomitant exercise of review or audit.

2.1 Case of a stand-alone intervention

2.1.1 Ex ante evaluation

An ex ante evaluation is performed before adopting or implementing the intervention. It gives support to the intervention design. It aims at having a direct influence on the decisions upstream from the implementation. It transposes lessons from past experiences into the Boxwork of the new intervention.

2.1.2 Mid-term of final

An evaluation during or at the end of the implementation is intended to draw lessons from the first years of the implementation and to adjust the contents of the ongoing intervention in relation to realities in the field and/or contextual developments. It often includes a report on outputs and an analysis of the first results and impacts achieved. It aims at improving the intervention under way and its conclusions may be supported by observations in the field.

2.1.3 Ex post

The ex post evaluation is performed right after or a long time after completion of the intervention. It is mainly concerned with assessing achieved impacts, identifying and judging unexpected impacts and verifying the sustainability of the intervention's benefits.

It enables to detect actual changes in the field and. If such changes occur soon enough, they can be analysed in order to arrive at an estimate of what is attributable to the intervention.

The ex post evaluation often aims to report to the institutions that have allocated the resources. Likewise, it helps to transfer lessons learned to other countries or sectors.

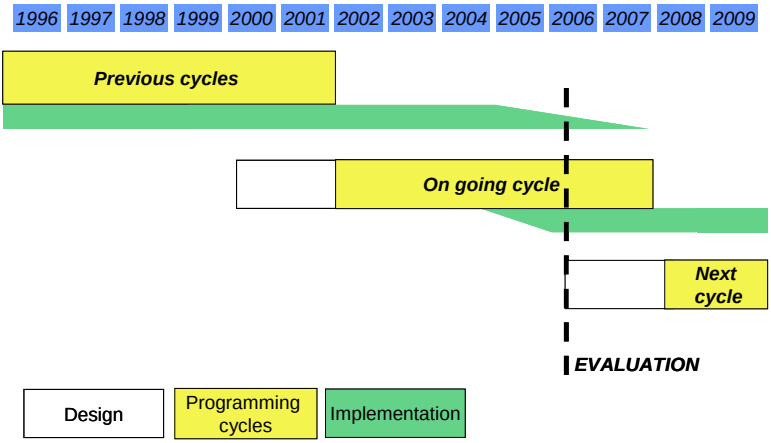
2.2 Case of a series of interventions

Many interventions are characterised by successive cycles and show a relative continuity between cycles. A new cycle typically starts before the intervention of the previous cycle has yielded all its effects.

In light of this, several cycles need to be considered when choosing the timing of the evaluation, in order to:

- Draw conclusions from an assessment of the effects of the intervention in the previous cycle
- Feed into the formulation of the intervention in the next cycle
- Observe the implementation of on going cycle and rapidly provide feedback.

Box 3 – Evaluation concerning successive cycles



3 Use of the evaluation

SUMMARY

Evaluation users

Evaluations are targeted at a variety of users, such as policy makers and intervention designers, managers and operators in charge of the implementation, partners, institutions having provided financing and to whom accountability is required, public authorities conducting connected or similar interventions, civil society organisation, and experts.

Types of use

An evaluation may help to make decisions, to articulate judgments and/or to better know and understand the intervention and its effects.

Dissemination of the evaluation

The dissemination process aims to make the final evaluation report known, but it also includes every other way in which findings, conclusions, lessons learned and recommendations can be drawn to users' attention. The dissemination process targets the Commission services, external partners, expert networks, the media and the wider public.

3.1 The evaluation users

Identifying the evaluation users is of utmost importance to optimise the usefulness of the evaluation for the various players, and especially to ensure that the evaluation:

- Meets the expectations of the intended users, in a way and at a time that fulfils their needs
- Has the required credibility vis-à-vis the intended users.

3.1.1 Policy makers and designers

Policy-makers and designers use the evaluation to prepare new interventions, the reform of existing interventions, the choice of strategic orientations, and decisions on allocation of budgetary, human and other resources, etc.

They need information that:

- Directly feeds into the decision-making process
- Arrives on time
- Answers their questions clearly, concisely, and reliably.

They are interested in strategic issues, external coherence and global impacts, which constitute the ultimate goal of the intervention.

3.1.2 Managers, partners and operators

The managers are responsible for the implementation and monitoring of the intervention, from headquarters to the field. The operators are the implementation agents closest to the public. Field level operators may either belong to The European Commission or to partner organisations sharing the responsibility of implementation.

They need information that arrives as early as possible to remedy problems or validate innovations. They are able to interpret complex and technical messages.

They are interested in the direct results of the intervention, in the needs and behaviour of targeted groups, and in interactions between partners.

3.1.3 Other actors

The institutions that funded an intervention expect accountability. This applies to the Parliament or the Council of Ministers, but also to all the co-funders. The taxpayers and citizens are also addressees of an evaluation.

The public authorities that conduct related or similar interventions are potential users of the evaluation, especially in the form of transfer of lessons learned. The same applies to the expert networks in the area of the intervention.

Finally, an evaluation is likely to be used by civil society actors, especially those representing the interests of the target groups.

3.2 Types of use

An evaluation may help:

- To prepare decisions

- To formulate value judgments and inform public opinion
- And/or to better know and understand the intervention and its effects.

The first type of use encompasses the second one (judging is a prerequisite to making decisions), which itself encompasses the third one (understanding is a prerequisite to formulating value judgements).

3.2.1 Assisting decision-making

The evaluation may be conducted for the benefit of those who have to decide upon or to negotiate about a change in the evaluated intervention. In such cases, it is used to adjust the implementation, to modify the intervention strategy or to redefine political orientations.

The evaluation may assist decision-making through two different channels:

- By formulating conclusions independently, and then proposing recommendations
- By favouring the involvement of the decision-makers concerned, or at least their close collaborators, with a view to encouraging take-up or direct feedback during the process.

Evaluation may assist decision-making in different ways, depending on the context of the decision:

- Recommendations may target the managers and/or designers of the intervention
- Recommendations may be addressed to all the partners and co-funders of the intervention
- Finally, an evaluation may be conceived as an aid to negotiation and problem-solving between a wider range of stakeholders , including interest groups and actors in civil society.

3.2.2 Assisting the formulation of value judgements

The evaluation may help users to shape their opinion on the merits of the intervention.

For instance, the formulation of an overall assessment is particularly useful for accountability purposes. In this case, the evaluation examines the merits of the intervention in relation to various viewpoints. It answers questions that are important for the funding institutions. The report is accessible to the general public. The independence of the evaluation and the transparency of the judgement are highlighted.

In this instance, particular attention is paid to the definition of judgement criteria (also called "reasoned assessment criteria"). The value judgements themselves are delivered only when the final report is submitted and its conclusions are discussed. Using the evaluation for accountability purposes therefore means having to wait for the end of the process.

3.2.3 Knowing and understanding

Apart from assisting in making decisions and formulating value judgements, which are the two main forms of use, the evaluation may also enable users to learn about the intervention, to better understand what works and what does not, and to accumulate knowledge. Indirectly, it contributes to transferring knowledge thereby acquired, to the benefit of professional networks that may not have a direct link with the evaluated intervention.

Unlike feedback, which directly concerns those responsible for the evaluated intervention, the transfer of lessons is an indirect process that takes place through networks of experts both within and outside the European Commission.

Such an accumulation of knowledge often starts during the evaluation process, through the experts who belong to the evaluation team or reference group. However, the transfer of lessons learnt may only occur after the final report has been delivered. A key step in this perspective is presentation of the evaluation in specialised networks, in the form of seminars or technical articles.

3.3 The dissemination of the evaluation

Dissemination concerns the final evaluation report, as well as all other means of publicising the conclusions, the lessons learned and the recommendations. Dissemination activities target the services of the Commission, European Institutions, external partners, networks of experts, the media, and the wider public.

3.3.1 Measures to be taken

Dissemination is planned when drafting the terms of reference, especially by specifying how the report will be published and what the evaluation team's role will be in that phase.

Throughout the process, all players should pay attention to the quality of the evaluation and its products (report, summary, annexes). The quality should be formally assessed at the end of the process.

The last evaluation reference group meeting enables to identify the main messages delivered by the evaluation and the targeted audiences.

After approval of the report, the communication plan is finalised by choosing the messages to be highlighted and the most suitable information channels for each audience.

Finally, the players should co-operate in order to implement the dissemination plan, which implies that the tasks and responsibilities be shared between the evaluation manager, the evaluation team (check that this is part of its mission and is specifically mentioned in the terms of reference), and the members of the reference group.

3.3.2 Which channels for dissemination?

The evaluation report is disseminated on the Internet and is thus accessible to all audiences. More active dissemination efforts are also undertaken for specific audiences:

- A one-page summary is written specifically for the hierarchy of the service that managed the evaluation. It highlights the main conclusions and recommendations.
- The report and/or its executive summary are sent to the services concerned and to the partners.
- A summary is also published on other relevant Intranet sites, with a link to the report.
- A summary is published on the Internet intended for the international development aid community. It highlights the lessons learned if they are transferable.
- One or more articles may be written for the general public or specialised networks.

- Finally, the report may be presented in meetings, workshops or seminars.

Practical advice for a presentation

A short presentation (10 to 20 minutes) is enough for delivering the main messages but more time needs to be left for questions (20 to 40 minutes).

The presentation covers the following points:

- The evaluated intervention
- The evaluation method
- The messages resulting from the evaluation
- Strengths and weaknesses of the messages.

On the website: examples of summaries and articles about an evaluation

4 The players and their roles

SUMMARY

Evaluation manager

The evaluation manager is a member of the service in charge of the evaluation. He manages the whole process on behalf of the European Commission.

Reference group

The reference group is presided over by the evaluation manager and provides assistance to the latter in monitoring and supervising the evaluation. It is composed of members of the European Commission's services and representatives from the partner countries whenever that is possible.

Evaluation team

The evaluation team is responsible for data collection and analysis as well as for the formulation of value judgements in response to the evaluation questions. This team writes and is responsible for the evaluation report. It submits its works regularly to the reference group and to the evaluation manager, and takes their comments into account.

Stakeholders

The stakeholders are individuals, groups or organisations that have a direct or indirect interest in the evaluated intervention and in the evaluation itself. They may be affected by the intervention or not.

4.1 Evaluation manager

The evaluation manager is a member of the service in charge of the evaluation. He manages the whole process from the beginning to the end on behalf of the commissioning body.

He is appointed:

- To ensure consistency throughout the evaluation process, from the terms of reference to the dissemination of the report and the follow-up of recommendations.

- To be the contact person for administrative issues and to coordinate the activity of the different actors (reference group and evaluation team).
- To organise, supervise and assess the quality of the different phases of the work.
- To ensure the smooth running of the evaluation.

Check-list: role of the manager

The evaluation manager

- forms the reference group after consulting the heads of the services concerned
- drafts the terms of reference after consulting the reference group members
- engages the evaluation team after consulting the reference group
- helps the evaluation team to get access to information and informants
- organises the discussion and approval of conclusions and recommendations in interaction with the reference group and the evaluation team
- performs a methodological quality assessment of the reports, double checked by a second person
- disseminates the evaluation to the different actors concerned
- monitors whether the recommendations are taken into account in interaction with the authorities concerned.

It is recommended to appoint a deputy to help the evaluation manager on a daily basis and to stand in when he is absent.

4.2 Reference group

The reference group is an interface between the evaluation manager and the evaluation team. It allows for the variety of viewpoints to be taken at the evaluated intervention.

Check-list: role of the reference group

Reference group members:

- Discuss and comment on the terms of reference drawn up by the evaluation manager
- Aggregate and summarise the views of the Commission services and act as an interface between the evaluation team and the services, thereby supplementing bilateral contacts
- Ensure that the evaluation team has access to and consults all information sources and documentation on activities undertaken
- Validate the evaluation questions
- Discuss and comment on notes and reports produced by the evaluation team. Comments by individual members of the reference group are compiled by the evaluation manager and subsequently transmitted to the evaluation team
- Assist in feedback of the findings, conclusions and recommendations from the evaluation.

If the evaluation is managed at headquarters, membership includes:

- Services of the Commission concerned by the intervention, including the Delegation concerned in case of a country/region level evaluation
- Specialists on a sector or a cross-cutting issue within the Commission
- Embassy of the partner country in the case of a country level evaluation.

If the evaluation is managed in the partner country, membership may extend to:

- Government services
- Selected development partners
- Selected civil society organisations
- Experts.

Experience has shown that it is preferable for the group to be limited in size (10 to 12 members) if it is to function effectively.

After identifying the services, institutions or organisations to invite to the reference group, the head of the service managing the evaluation sends them an invitation to:

- Officially announce the launching of the evaluation
- Request that they appoint one person as a member of the reference group.

Box 4 – Timetable and role of the reference group

Stage	Meeting	Role
Terms of reference		Comments
Engagement of the evaluation team		Comments
Evaluation questions	1	Validation
Inception report		Comments
Desk report	2	Discussion Comments
Field phase debriefing	3	Discussion
Version 1 of the final report, conclusions and recommendations	4	Discussion Comments

On the website: check-list for organising and moderating a reference group meeting

4.3 Evaluation team

The evaluation team is responsible for data collection and analyses and for the evaluation report, including the formulation of value judgements (also called reasoned assessment) and the drafting of conclusions and recommendations. It interacts with the reference group and the evaluation manager. It provides the commissioning body with evaluation services in the Boxwork of a contract.

The members of the evaluation team are independent from the organisations that participated in the design and implementation of the evaluated intervention.

Criteria for selecting an evaluation team

- Knowledge of the evaluated development aid modalities
- Knowledge of the evaluation scope (e.g. sector, country)
- Presence of local experts and consultants
- Team leader's managerial skills (organisation, coordination, dialogue with the client, etc.)
- Team's independence as regards the evaluated intervention and absence of conflicting interests
- Evaluation skills and particularly the ability to formulate value judgements, to draw up conclusions and recommendations and to draft syntheses
- Mastery of data collection and analysis tools
- Diversity of team members' profiles and expertise
- Price.

The evaluation manager sets the relative weight of criteria and specifies it in the terms of reference.

Involve local consultants

Involving local consultants, as far as possible, helps to promote the development of local capacity (development of national evaluation expertise) and to benefit from their close knowledge of the field.

4.4 Stakeholders

The stakeholders are individuals, groups or organisations that have responsibilities and/or direct or indirect interests in the evaluated intervention. They may be affected by the intervention or not.

Involving stakeholders in the evaluation is useful in order to:

- Take their points of view into consideration
- Benefit from their expertise and knowledge
- Encourage them to use the evaluation.

4.5 Quality assurance

The evaluation manager has to assure that the conclusions build upon a strong factual and methodological basis.

Within the Evaluation Unit at EC headquarters, managers have the training and skills enabling them to assess the methodological quality of evaluation reports. All evaluations are monitored by two members of the Unit, both of whom participate in the quality assessment.

In the case of devolved evaluations, the manager uses the quality criteria proposed by these guidelines as well as the suggestions given to assess them, if necessary with the help of the Evaluation Unit.

On the website: quality assessment grid and suggestions for filling the grid

III. ON METHODS

The following chapters are dedicated to eight key components of the evaluation methods presented in a logical way (and not a chronological one): intervention strategy, evaluation questions (usefulness, feasibility, preparation), judgement references (criteria and indicators), methodological design, data collection, analysis, value judgement (conclusions, lessons and recommendations) and quality assurance.

5 Intervention strategy

SUMMARY

The first step of the evaluation process is a review of the intervention strategy, that is its rationale, its logic and its connections with other policies.

Intervention rationale

The rationale of an intervention is to satisfy needs, solve problems or tackle challenges that are considered to be priorities in a particular context and that cannot be addressed more effectively in another way. During the evaluation the main points of the rationale should be highlighted or reformulated if the programming documents lack precision.

Intervention logic

The intervention logic identifies the activities, outputs, results and different levels of expected impacts. It is necessary to restate or reconstruct the intervention logic in order to understand which effects are expected and therefore which evaluation questions are to be asked.

Related policies

It is necessary to identify other related policies (those of European institutions, member States, other donors and partner countries). The main related policies are identified in order to understand where complementarities, potential synergies, risks of duplication, and coordination needs lie.

5.1 Intervention rationale

The rationale of an intervention is to satisfy the needs, solve the problems or tackle the challenges that are considered to be priorities in a particular context and that cannot be addressed more effectively in another way.

In principle, the intervention rationale is justified in the programming documents. During the evaluation the main points of the rationale should be highlighted or reformulated if the programming documents lack precision.

The evaluation identifies the reasons for which the priorities have been chosen, for example: priorities of the policy in which the intervention takes place, urgency of the needs to be satisfied, advantages of the intervention as compared to alternative options.

The analysis proceeds as follow:

Examination of problems and answers

The evaluation report succinctly sets out the following:

- Context of the intervention when it was initiated
- Main problems diagnosed (needs, challenges)
- Why certain strategic options have been chosen rather than others.

One or several evaluation questions may focus on the problems addressed by the intervention, for example:

- Are there alternative options for solving the problems identified?
- Has the nature of the problem which justified the intervention changed?
- What is the precise extent of the needs justifying the intervention?

Examining the rationale

Where relevant, the evaluation report highlights the following:

- Justification of the fact that the needs, problems and challenges cannot be addressed by private initiatives.
- Justification of the fact that they cannot be addressed more effectively by other public initiatives.

Recommendations

The official documents often focus on the strategic options that were finally selected. If the evaluation team wants to know what the alternatives were, it may be useful to hold interviews with key informants, whilst remaining wary of possible biases.

5.2 Reconstructing the intervention logic

The intervention logic identifies all the activities and expected effects (outputs, results and impacts) of an intervention (see Box 8), as well as the assumptions that explain how the activities will lead to the effects in the context of the intervention.

Reconstructing the intervention logic is necessary:

- To help clarifying the objectives and translating them into a hierarchy of expected effects so that they can be evaluated
- To suggest evaluation questions about these effects
- To help assessing the internal coherence of the intervention.

The intervention logic may be "faithful" to the programming documents and to the documents establishing the policy in which the intervention takes place. In this case, the expected effects are inferred from the stated objectives in the official documents.

When the intervention logic is reconstructed during the evaluation, implicitly expected effects that were not mentioned in the initial documents may be taken into account. The fact that this is no longer a "faithful logic" must then be mentioned. The "faithful" approach is relevant when the objectives are expressed precisely and in a verifiable way. The other option is preferable if objectives are too vague or ambiguous.

The intervention logic often evolves over time. In such cases, it is justified to reconstruct it for successive time periods.

The intervention logic is a useful simplification of reality, but one has to bear in mind that the real world is complex. In addition to the reconstruction of the intervention logic, it is useful to identify the main external factors that enable or constrain the implementation and the effects. One also has to remember that actual cause-and-effect explanations are often more complex than initial assumptions.

5.2.1 How can it be reconstructed?

Reconstructing the intervention logic proceeds as follow:

- Collect and analyse the official documents establishing the intervention and allocating resources
- Identify the main activities
- Identify the objectives
- Translate the objectives into expected results and impacts.
- Connect the activities to the expected impacts by reconstructing the cause-and-effect linkages
- Check that cause-and-effect linkages are logical, i.e. considered as plausible in the light of available knowledge
- Discuss the reconstructed logic with a few key informants (designers and managers) and with experts of the concerned policy domain/country
- Present and discuss the intervention logic in a reference group meeting.

5.2.2 The most common presentations

Logical Boxwork

This technique consists in a matrix presentation which specifies the objectives (first column), how their achievement should be verified (second column: indicators, and third column: sources), and the main assumptions on the external factors (fourth column).

The lines of the logical Boxwork are a simplified presentation of the intervention logic, on only four levels: activities, results, purpose (i.e. direct benefit for the targeted group) and goal.

This representation is adequate for a simple intervention, like a project or a homogeneous programme. However, it cannot fully grasp the complexity of a heterogeneous intervention such as an integrated rural development programme, a country strategy or a global sector policy.

Due to its simplified nature, the logical Boxwork allows for specifying the indicators that should be used at each level, as well as most external factors.

Diagram of objectives

This technique (see Volume 4) consists of an identification of officially stated objectives and a graphical presentation of the cause-and-effect linkages between objectives, from the most operational to the most global. The intervention logic is represented in the form of boxes and arrows.

A particular form of representation is the objectives tree. It is applied in the case where each objective on a lower rank contributes towards a single objective on a higher rank.

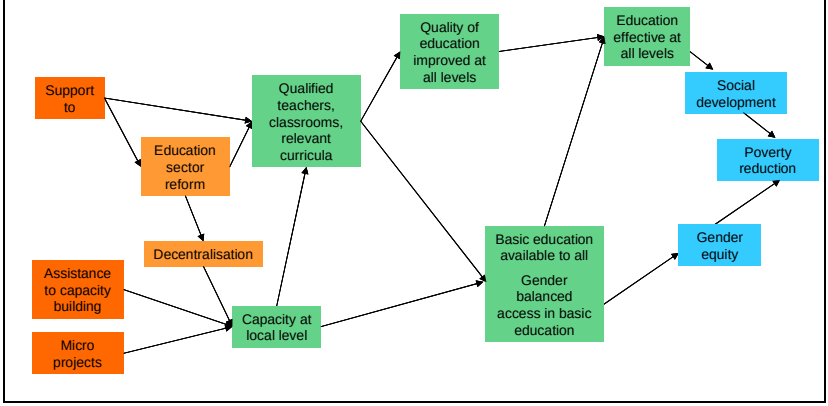
Diagram of expected effects

This technique (see Volume 4) is similar to the diagram of objectives since it also builds upon officially stated objectives. However, the objectives are translated into expected effects before being presented as a diagram. By translating objectives into expected effects, more concrete and easily verifiable concepts can be worked on.

Unlike the logical Boxwork, the diagram can have as long chains of objectives as necessary. It can be used to highlight synergies and intricate relations between objectives. This presentation is appropriate for complex interventions such as integrated programmes, country strategies or sector policies.

However the diagram does not allow illustrating external factors, conditions and risks as the fourth column of a logical Boxwork does. Those elements are specified at the methodological design stage (see Chapter 8) on the basis of a question-per-question analysis.

Box 5 – Diagram of expected effects



5.3 Related policies

The idea is to identify the main related interventions and to situate the evaluated intervention in relation to them.

The identification of related policies is useful:

- To help ask questions belonging to the coherence/complementarity and relevance families (see Section 6.4)
- If necessary, to examine the quality of the intervention design, and especially the fact that the objectives are complementary to those of the other related policies.

The evaluation report briefly sets out the following:

- Main national and international institutions active in the same region and sector and/or targeted at the same group
- Main policies implemented by these institutions in the same region, in the same sector and/or targeted at the same group.

Certain evaluation questions may concern related policies for a more in-depth examination, for example:

- Is the intervention consistent with and complementary to the policies and priorities of the partner country and/or other Community policies?
- Has the implementation been coordinated with the actions of the other sponsors and has the complementarity been improved?
- To what extent is there value added when the intervention is conducted at EC rather than member-State level?

6 Evaluation questions

SUMMARY

Usefulness of the questions

Evaluation questions focus the evaluation work on a limited number of key points, thus allowing more targeted data collection, more in-depth analysis and a more useful report.

Origin of the questions

Some evaluation questions are inferred directly or indirectly from the intervention logic. Other questions do not require a preliminary examination of the intervention logic because they concern effects that are not featured in it. They can also concern cross-cutting issues (gender, environment, etc.).

Selecting questions

The questions (10 maximum) are selected as regard the potential usefulness of answers and their feasibility. The set of questions has to be discussed at the inception meeting.

Questions and evaluation criteria

The questions are classified according to the evaluation criteria formalised by OECD-DAC (relevance, effectiveness, efficiency, sustainability, impact), plus two criteria that apply to all EC policies (coherence/complementarity, and Community value added).

Preparing an evaluation question

It is important to ensure that the answer to the question will be useful, to specify the nature of the expected utilisation, to ensure that the question pertains to evaluation, to specify the scope of the question, and to link the question to the intervention logic and to an evaluation criterion.

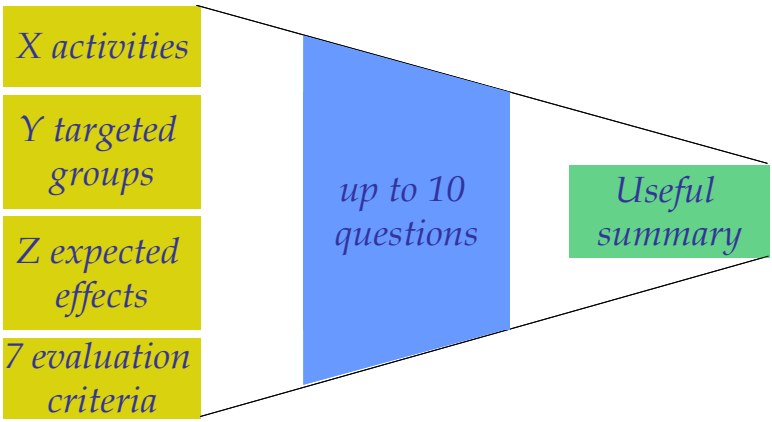
6.1 Usefulness of the questions

Evaluation questions focus the evaluation work on a limited number of key points, thus allowing more targeted data collection, more in-depth analysis and a more useful report.

Focusing an evaluation on several key points is particularly useful when the evaluated intervention is multidimensional (see Box 1). In that case, if one wished to study all the dimensions of aid through all evaluation criteria, the work would be extremely costly or else very superficial. Thus, choices have to be made.

As illustrated in the Box below, the choice of the questions contributes to optimise the usefulness of the evaluation.

Box 6 – Focus the evaluation



6.2 Origin of the questions

6.2.1 Questions inferred from the intervention logic (directly)

The intervention logic is the set of all the assumptions explaining how the intervention will meet its objectives and produce the expected effects (see Chapter 5).

Once the evaluation team has identified the expected effects and the cause-and-effect assumptions between them, it is possible to ask all types of question such as:

- To what extent has [activity X] contributed to [generating effect Z]?
- To what extent have [activities X1, X2, X3, etc...] contributed to [generating effect Z]?

- To what extent have [activities X1, X2, X3, etc...] contributed to [generating effects Z1, Z2, Z3, etc.]?

These questions are directly derived from the intervention logic in a standard form and belong to the effectiveness family (achieving one or more expected effects).

6.2.2 Questions inferred from the intervention logic (indirectly)

Each standard question can be reformulated in many ways:

... for instance by specifying the scope:

- To what extent has the intervention *[and more specifically the one implemented by means of instrument A or procedure B]* contributed towards generating effect Z?
- To what extent has *[coordination with the other development partners]* contributed towards generating effect Z?

... or by specifying the effect concerned:

- To what extent has the intervention contributed towards generating effect Z *[for the poorest population groups]*?

... or by changing the evaluation criterion (see Section 6.4):

- To what extent has the intervention generated effect Z *[with a strong probability of survival of effects after the end of the aid]*? (sustainability)
- To what extent has the intervention contributed towards generating effect Z *[at a low cost as compared to …]*?
- When the intervention targets effect Z, to what extent does this *[correspond to the needs of the population concerned]*? (relevance)
- When the intervention targets effect Z, to what extent is this *[compatible with or contrary to the objectives of other EU policies]*? (coherence/complementarity).
- When the intervention targets effect Z, to what extent does this *[add value to a similar intervention implemented by member States]*? (Community value added).

6.2.3 Other questions

The following questions do not require a preliminary examination of the intervention logic because they concern effects that are not featured in it:

- To what extent has [activity X, instrument A, procedure B] generated unexpected effects? If it has, who has benefited or lost out?
- To what extent has The European Commission integrated [cross-cutting issue X] into the design and implementation of its interventions?

6.3 Selecting questions

The choice of evaluation questions proceeds through the following steps:

Identify questions

A first version of the evaluation questions is proposed on the basis of:

- The analysis of the intervention logic
- The analysis of the intervention rationale
- Issues that justified the decision to launch the evaluation
- Issues to be studied, as stated in the terms of reference
- Questions raised in the ex ante evaluation, where relevant.

In a second version, the list and wording of the evaluation questions also take into account:

- Issues raised by key informants at the start of the evaluation
- Expectations of the members of the reference group
- Subjects raised by the evaluation team.

Assess the potential usefulness of answers

It is necessary to assess the potential usefulness of the answers (assuming that the questions will be properly answered) by considering the following points:

- Who is to use the answer?
- What is the expected use: knowledge, negotiation, decision-making, communication?

- Will the answer arrive in time to be used?
- Is the answer not already known?
- Is there not another study (audit, review) underway, likely to provide the answer?

If the set of questions has to be discussed in a meeting, it may be useful to classify them in three categories of potential utility: higher, medium, lower.

Check that nothing important has been overlooked

Experience has shown that it is most harmful to the quality of the evaluation if the following types of questions are left out:

- Questions on efficiency and sustainability
- Questions concerning negative effects, especially if those effects concern underprivileged groups
- Questions concerning very long-term effects.

Assess the feasibility of questions

The feasibility (evaluability) of a question should be examined, but always after its usefulness. For this purpose the following actors should be consulted:

- The service managing the intervention
- One or more experts in the field
- One or more evaluation professionals.

If a question is potentially very useful but difficult to answer, check whether a similar question would not be easier and equally useful. For example, if a question concerns a relatively far-reaching or global impact, its feasibility could probably be enhanced by focusing on the immediately preceding impact in the intervention logic.

The most frequent limitations to evaluability

- **Managerial weaknesses**

The monitoring data and management reports are inadequate or unreliable.

Managerial difficulties have generated conflicts that restrict the access to certain informants or cause those informants to bias their answers.

In case of a strong suspicion of illicit or illegal practices, it is preferable to postpone the evaluation question for later and to start with an audit.

- **Too complex scope**

In view of the available time and budget, there are too many data to collect, informants to meet and analyses to perform, and they are too dispersed.

- **Too recent activity**

The activity has not yet produced observable effects.

The informants have not yet stabilised their opinions.

- **Highly innovative activity**

It is difficult to define the terms of the question without ambiguity.

There is a lack of expertise to understand the cause-and-effect mechanisms.

- **Too marginal intervention**

The evaluated activity does not attain the critical mass that would allow an analysis of its contribution.

- **Far-reaching impact**

There are so many external factors and they are so influential that it is impossible to analyse the contribution of the intervention.

Discuss the choice of key questions

The set of questions is discussed at the inception meeting.

The selection is more likely to be successful if potential users have been consulted and have agreed upon the selected questions, and if no legitimate view has been censored.

Reasons for selecting a question

- **Because someone raised it**

Those who submit a question tend to cooperate in answering it and to accept the answer.

An actor may ask a question primarily with the intention of influencing or even obstructing the action of another actor. The potential usefulness of this type of question has to be examined carefully.

- **Because it is useful**

The intervention or one of its aspects is innovative and several actors expect a validation.

A decision is going to be taken and the conclusions may arrive in time to help in taking that decision.

A public debate is planned and the conclusions may be delivered in time to feed into the debate.

- **Because the answer is not known**

The question has not been already answered by another evaluation, an audit or a study.

The question has not been already asked in many other evaluations with always the same answer.

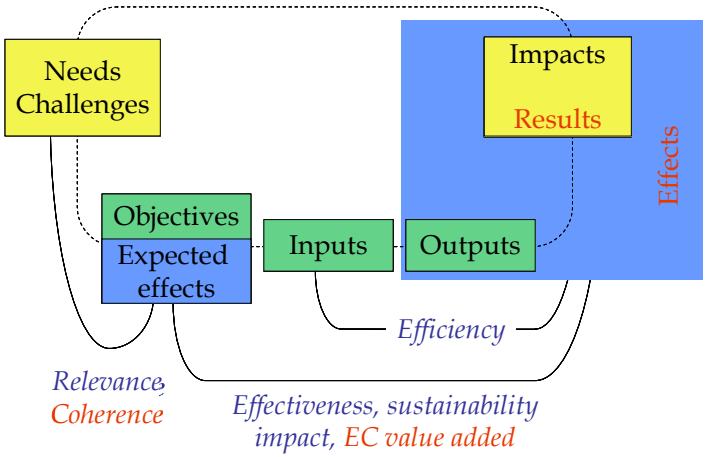
On the website: examples of questions for a country/region evaluation, for a global, thematic or sectoral evaluation

6.4 Questions and evaluation criteria

The questions are classified in different families that correspond to different "viewpoints" on what is being evaluated. Seven of these viewpoints, also called evaluation criteria, are to be considered: relevance, effectiveness, efficiency, sustainability, impact, coherence/complementarity, and Community value added.

The first five correspond to the traditional practice of evaluation of development aid. They have been formalised by the OECD (DAC). The following two apply to all EC policies.

Box 7 – Evaluation criteria



Relevance

The extent to which the objectives of the development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and EC's policies.

- Example: To what extent does the concentration of aid on basic education correspond to the needs of the partner country?

Effectiveness

The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.

- Example: To what extent has the aid contributed to equal access to high-quality basic education?

Efficiency

The extent to which outputs and/or the desired effects are achieved with the lowest possible use of resources/inputs (funds, expertise, time, administrative costs, etc.)

- Example: Has implementation in the form of sector-specific financial aid made it possible to obtain the same effects

with lower transaction costs for The European Commission and the partner country?

Sustainability

The continuation of benefits from a development intervention after major development assistance has been completed.

The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.

- Example: To what extent has the aid contributed towards durably remedying the backlog in road network maintenance?

Impact

Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

- Example: From the point of view of the groups concerned, are environmental nuisances acceptable compared to the positive effects of the intervention?

Coherence/complementarity

This criterion may have several dimensions:

Coherence within the Commission's development programme

- Example: Can it be said that the activities and outputs logically allow the objectives to be achieved? Are there contradictions between the different levels of objective? Are there duplications between the activities?

Coherence/complementarity with the partner country's policies and with other donors' interventions

- Example: Can it be said that there is no overlap between the intervention considered and other interventions in the partner country and/or other donors' interventions, particularly Member States?

Coherence/complementarity with the other Community policies

- Example: Is there convergence between the objectives of the intervention and those of the other Community policies (trade, agriculture, fishing, etc.)?

Community value added

The extent to which the development intervention adds benefits to what would have resulted from Member States' interventions only in the partner country.

- Example: To what extent has the sharing of roles between The European Commission and Member States contributed to optimise the impact of the support?

6.5 Preparing an evaluation question

6.5.1 Ensure that the question pertains to evaluation

Before drafting a question, ensure that it does not concern audit or monitoring

If the question concerns only compliance with rules and procedures, it is an audit rather than an evaluation question.

If the question covers only the completion of outputs, it is a monitoring question.

6.5.2 Specify the scope of the question

The scope of the question can be the entire intervention or a particular dimension of its design or implementation, for instance:

- Has the intervention helped to generate effect Z as expected? (entire intervention)
- Has the stakeholders' participation in the formulation of the intervention helped to generate effect Z more successfully (design modality)
- Have the measures taken to ensure coordination with the other donors helped to generate effect Z more successfully? (implementation modality)
- Has the choice of sector budget support helped to generate effect Z more successfully (funding modality).

6.5.3 Link the question to the intervention logic

Which effect does the question concern? For example:

- Has the intervention helped to generate effect Z as expected?

- Has the intervention helped to generate effect Z for the poorest groups?
- Has the intervention increased the poorest groups' chances of obtaining effect Z?

If the question concerns a short-term result or specific impact, it will be more easily evaluable than if it concerned a global impact. Conversely, a question on a global impact will be more useful for decision-makers at a strategic level.

6.5.4 Link the question to an evaluation criterion

The question is drafted with an evaluation criterion in mind (e.g. relevance, effectiveness, efficiency, sustainability, impact, Community value added, and coherence/complementarity) (see Section 6.4), unless it is designed only to further knowledge and understanding.

6.5.5 Finalise the writing of the question

It is better to choose open wording that calls for a qualified answer than a closed question that requires a "yes or no" answer. Furthermore, the drafting of the question should be simple and concise. For information purposes, it is useful to add comments on all or some of the following points:

- Details on the scope of the question
- Definition of the main terms used.

On the website: examples of questions for different categories of evaluations

7 Judgement references

Summary

Judgement criteria

A judgement criterion (also called "reasoned assessment criterion") specifies an aspect of the merits or success of the evaluated intervention. It is used to answer an evaluation question positively or negatively.

Target levels

To formulate an explicit judgement, it is recommended to refer both to a judgement criterion and to a target, that is, to a performance standard, a success threshold, a benchmark or a good practice as a point of comparison.

Indicators

Indicators specify in advance which data are to be collected. They thus help to focus the data collection process.

The evaluation focuses primarily on result and impact indicators used in the Boxwork of surveys, management databases or statistical sources

From question to criterion and to indicator

This chapter proposes an approach for switching from evaluation questions to judgement criteria and indicators.

7.1 Judgement criteria

A judgement criterion specifies an aspect of the evaluated intervention that will allow its merits or success to be assessed. Whilst "judgement criterion" is the appropriate word, an acceptable alternative is "reasoned assessment criterion".

Be careful not to confuse concepts!

Judgement criteria presented in this chapter.

The evaluation criteria: relevance, effectiveness, efficiency, sustainability, impact, Community value added, and coherence/complementarity (see Section 6.4).

The quality criteria applying to evaluation reports (see Section 12.1).

The following is an example of a question:

- To what extent has EC support improved the capacity of the primary education system to enrol pupils from underprivileged groups without discrimination?

The question belongs to the effectiveness family.

Like most evaluative questions, it has two parts:

- What is being judged: "EC support".
- The way of judging: Has it "… improved the capacity of the primary education system to enrol pupils from underprivileged groups without discrimination".

The judgement criteria are meant to develop and specify the second part of the question, for example:

- Criteria 1: Capacity of the primary education system to enrol pupils from ethnic minority E satisfactorily.
- Criteria 2: Capacity of the primary education system to enrol pupils from disadvantaged urban areas satisfactorily.

To be used in practice, each judgement criterion has to be accompanied by a target level and one or more indicator(s).

Recommendations

Always define the judgement criterion before selecting an existing indicator or designing a new indicator. It is essential to clarify the concepts first. By focusing too early on indicators, one is likely to get trapped into existing information, even if it is inadequate for answering the question asked.

Have the definition of the judgement criteria discussed by the reference group so that the diversity of opinions relevant to the intervention can be taken into account.

There may be disagreement on the judgement criteria, for instance the same effect may have a dimension that is judged positively by certain members of the reference group and another dimension judged negatively by others. In this case there are two options: (1) choose only one judgement criterion but be careful to avoid biasing the evaluation; or (2) choose several criteria, although this will increase and complicate the data collection and analysis work.

To optimise the collection and analysis of data, it is best to define a limited number of judgement criteria for each question. This recommendation also takes into account users' capacity to absorb information.

Where relevant, mention any gap between the criteria used to formulate the value judgements at the end of the evaluation process and those selected in the first phase (desk) of the evaluation.

7.2 Target levels

The concept of a « target » is widely used in the context of public management for setting a verifiable objective or a level of performance to be achieved. In an evaluation context it is used in a much wider sense since the evaluated intervention may have to be judged against targets that were set in advance or that are specifically identified at the time of the evaluation.

7.2.1 Objective defined in a verifiable way

This type of target is derived from one of the intervention objectives, provided that it has been set in a verifiable way. In this particular case, the same indicator helps to define the objective, to make the judgment criterion operational and to determine the target.

- Example: the number of qualified and experienced teachers per 1,000 children of primary-school age should reach at least 20.

7.2.2 Comparable good practices

This type of target is established at the outset of the evaluation. It is not related to an objective or a performance Boxwork existing prior to the evaluation.

- Example: the access to primary education with qualified and experienced teachers is at least as satisfactory as in the case of X (recognised good practice at regional level).

The procedure is as follows:

- Identify a comparable practice recognised for its quality (similar EC intervention in another country; intervention by another donor, intervention in another sector though using the same instruments)
- Obtain information on the practice to be compared (this is easier if it has already been evaluated)
- Ensure that the contextual conditions are close enough to allow for comparison
- Proceed with the comparison (most often qualitative)
- Discuss and validate the comparison with the reference group.

7.2.3 Best practices identified within the intervention

This type of target is found within the evaluated intervention itself, during the synthesis phase, provided that specific practices can be considered as good as regards the judgement criterion under consideration.

In this case, the good practice will be used as a benchmark to judge the other ones. Of course, it is advisable to check that the

contextual conditions are close enough so as to allow for comparison.

- Example: In areas where ethnic minority X concentrates, the number of qualified and experienced teachers per 1,000 children of primary-school age is close to the best performing areas in the country.

7.3 Indicators

The evaluation team may use indicators in order to:

- Collect and process data in a form that can be used directly when answering questions
- Avoid collecting an excessive amount of irrelevant data and to focus the process on the questions asked.

A quantitative indicator is based on a counting process (e.g. number of qualified and experienced teachers). The basic indicator directly results from the counting process. It may be used for computing more elaborate indicators (ratios, rates) such as cost per pupil or number of qualified and experienced teachers per 1,000 children of primary-school age.

A qualitative indicator (or descriptor) takes the form of a statement that has to be verified during the data collection (e.g. parents' opinion is that their children have the possibility of attending a primary school class with a qualified and experienced teacher).

> **Warning!**
>
> Unstructured data are also collected during the evaluation, either incidentally, or because exploratory tools are used, such as case studies. This kind of evidence may be sound enough to serve a basis for deriving conclusions, but it is not an indicator.

7.3.1 Evaluation indicators

The main evaluation indicators are those related to the judgement criteria. They specify the data needed in order to formulate a judgement based on those criteria.

An indicator can be constructed specifically for an evaluation (ad hoc indicator) and measured through a survey. It may also be

drawn from a monitoring database, from a performance assessment Boxwork, or from statistical sources.

Evaluation indicators and others

When an evaluation question pertains to an intended result or impact, it is worth checking whether this result or impact has been subject to performance monitoring. In such cases, the evaluation team uses the corresponding indicators and data, which is a considerable help, especially if baseline data have been recorded.

Performance monitoring may, however, be of little or no help in the instance of evaluation questions relating to cross-cutting issues, sustainability factors, unintended effects, evolving needs or problems, coherence, etc.

7.3.2 Categories of indicators

Indicators and the intervention cycle

Indicators are used throughout the intervention cycle. They are first used to analyse the context; then, for the choice and validation of the intervention strategy, afterwards for monitoring outputs and results and, finally, for the evaluation.

Indicators, monitoring and performance assessment

Monitoring systems and performance assessment Boxworks use indicators which derive from the diagram of expected effects (also called results chain).

Monitoring indicators primarily relate to inputs and outputs. Performance indicators primarily focus on intended results.

Indicators and evaluation

Evaluation indicators are used to help answering specific evaluation questions. Depending on the question, they may relate to the needs, problems and challenges which have justified the intervention, or to the achievement of intended outputs, results and impacts.

Global and specific indicators

Global or contextual indicators apply to an entire territory, population or group, without any distinction between those who

have been reached by the intervention and those who have not. They are mainly taken from statistical data. A chapter of Volume 4 offers help to look for contextual indicators.

Specific indicators concern only a group or territory that has actually been reached. Specific indicators track the changes among those affected by the intervention. Most of specific indicators are quantified or qualified through surveys and management databases.

On the website: how to build up a scoring grid?

Indicators and intervention logic

Indicators may be used to support the identification of the resources, realisations, results or impacts.

Box 8 – Indicators and intervention logic

Category	Definition
Input indicators	Financial, human, material, organisational or regulatory resources mobilised during the implementation of the intervention.
Activity indicators	Implementation and management processes.
Output indicators	Goods and services that are delivered under the responsibility of the managers of the intervention. To put it simply, it can be said that outputs are what is bought with the public money.
Result indicators	Immediate effects of the intervention for its direct addressees. An effect is immediate if the operator notices it easily while he is in contact with the addressees. Because they are easily acknowledged by the operators, result indicators can be monitored systematically.
Impact indicators	Far reaching and indirect consequences of the intervention.

7.4 From question to indicator

The Box 9 proposes an example of switching from an evaluation question to the corresponding indicator. It should be kept in mind that several indicators may be derived from a judgement criterion, and that several judgement criteria may be derived from a question.

In the example below, the question refers to a family or evaluation criteria: the effectiveness.

The judgement criterion derives from the question as following:

- It specifies the type of success to be assessed by asking the question
- It specifies several concepts such as "education system ", "disadvantaged groups" and "discrimination".

Box 9 – From question to indicator: example

Question	To what extent has EC support improved the capacity of the education system to enrol pupils from disadvantaged groups without discrimination?
Judgement criterion/criteria	Capacity of the primary education system to enrol pupils from ethnic minority X with satisfactory quality.
Indicator(s)	Number of qualified and experienced teachers per 1000 children of primary-school age in areas where ethnic minority X concentrates.

The indicator derives from the judgement criterion as following:

- It describes in detail the information required to answer the question according to the judgement criterion chosen
- The indicator derives from the judgement criterion ("capacity of the primary school system to enrol pupils with satisfactory quality")
- It is quantitative, but a qualitative indicator could also have been defined, for instance: « Surveyed parents

confirm that their children have the possibility of attending a primary-school class and benefit from a qualified and experienced teacher".

7.4.1 Quality of an indicator

An indicator measures or qualifies with precision a judgement criterion or a variable under observation (construct validity). If necessary, several less precise indicators (proxies) may be used to enhance validity.

An indicator provides straightforward information that is easy to communicate and is understood in the same way by the information supplier and the user. It is precise, that is, associated with a definition containing no ambiguity. It is sensitive, that is, it generates data which vary significantly when a change occurs in what is being observed.

Performance indicators and targets are often expected to be SMART, i.e. Specific, Measurable, Attainable, Realistic and Timely. The quality of an evaluation indicator is assessed differently.

Indicators and effects: a warning!

An indicator used to evaluate an effect is not in itself a measurement or evidence of that effect. The indicator only informs about changes, which may either result from the intervention (effect) or from other causes.

The evaluation team always has to analyse or interpret the indicator in order to assess the effect (see Chapter 10).

8 Methodological design

SUMMARY

Designing the evaluation method consists in setting up the approach that will allow the evaluation team to answer the questions and to come to an overall assessment. The methodological design includes: a strategy for collecting and analysing data, a selection of investigation areas, a series of specifically designed tools and a work plan.

Design table per question

The design table explains how an evaluation question will be answered, including the chain of reasoning which connects data, findings and conclusions. It includes: the scope of the question, the judgment criteria, indicators, target levels, the chain of reasoning, the analysis strategy, the investigation areas, information sources and tools.

Chain of reasoning

The chain of reasoning describes how the evaluation team plans to answer the question. The purpose is not to complicate the question but to help giving an answer.

Optimising the overall design

The purpose is to finalise the overall evaluation method in a way which cuts across the evaluation questions and which makes a good enough mix of evaluation tools, considering the available time and resources.

8.1 Introduction

The methodological design consists in setting up the approach that will allow the evaluation team to answer the questions and to come to an overall assessment. It includes:

- The chain of reasoning that will be followed
- A strategy for collecting and analysing data
- Selected investigation areas
- A series of specifically designed tools
- A work plan.

The evaluation team starts designing a provisional method as early as from the drafting of its proposal in order to draw up cost estimates. A key assumption at this stage is the extent to which the evaluation will rely on secondary data or will involve specific data collection work in the field.

The main Box of the method is then established during the inception stage, in line with the evaluation questions, judgement criteria, indicators, data collection tools and analysis strategy.

The method is refined and finalised before the field phase and fully described in the desk phase report.

The final report includes a short and sharp presentation of the evaluation method, together with its limitations, if there are any. The method is fully described in annex, including initial design, problems encountered, solutions found, method actually implemented, and limitations.

The evaluation method is designed through an iterative process at three levels:

- A question-by-question approach, allowing the evaluation team to prepare the design tables with an aim to adequately answer each question
- An overall approach which cuts across the questions, and which allows the evaluation team to optimise the method as a whole, whilst matching time and resource constraints
- A series of specifically developed tools.

8.2 Design table per question

The design table explains how an evaluation question will be answered, including the chain of reasoning which connects data, findings and conclusions.

A design table is developed for each question.

Box 10 – Structure of a design table

Question	Text of the question
Comments	Why is the question asked?
Scope	What does the question cover?
Judgement criterion / criteria	How will the merits and success be assessed?
Indicator(s)	Which data will help assessing the merits and success?
Target level(s)	Which level or threshold can be considered as a success?
Chain of reasoning	Steps of reasoning planned to answer the question by: - quantifying / qualifying indicators - analysing information - formulating the value judgement
Analysis strategy	Type(s) of analysis to be applied (see Section 10.1.1)
Investigation areas	Areas where data are to be collected and analysed (see Section 8.3.2)
Information sources and tools	What will be the origin of data (see Section 8.3.3)

This table is progressively refined in successive versions:

- Preliminary version appended to the inception report
- Draft version(s) prepared during the desk phase of the evaluation as the methodological design is progressively optimised

- Final version attached to the desk phase report.

On the website: example of design table

8.2.1 Chain of reasoning

The chain of reasoning includes all the steps through which the evaluation team plans to answer the question. The purpose is not to complicate the question but to help giving an answer. The intended reasoning is indicative but it is worth clarifying it in advance because:

- The reference group members get an opportunity to provide specific advice and inputs
- All evaluation team members understand why they are collecting and analysing data, and therefore work more effectively
- The team members who are not familiar with evaluation receive useful guidance on which data are to be collected and analysed.

Steps pertaining to indicators

These steps may pertain to:

(1) the current level / status of the indicators, possibly with a break-down per country, area, social group, etc., for instance:

- What is the current value of quantitative indicator X at country level, and for targeted areas/groups?

(2) changes in the indicators, for instance:

- Do stakeholders perceive a positive change in qualitative indicator Y over the evaluated time period?

Steps pertaining to analysis

These steps are detailed with a view to:

(3) Confirming assumptions about the success of the intervention and substantiating a positive answer to the question, for instance:

- Has the design of EC support included a commitment to monitor performance indicators related to effect Z?
- Was such monitoring actually undertaken?
- Was the monitoring data subject to periodic discussion among development partners?

- Have partners taken action as a follow-up to such discussions?
- Were such actions designed with a view to achieving effect Z?
- Etc.

(4) Challenging assumptions about the success of the intervention and substantiating a negative answer to the question, for instance:

- Have other development partners pushed for monitoring performance indicators related to effect Z?
- Have non-state actors contributed to putting the issue of achieving effect Z onto the political agenda?
- How far did other partners contribute towards shaping the actions taken in favour of disadvantaged groups?

Steps pertaining to judgement criteria

These steps are meant to assist in the formulation of conclusions involving explicit value judgements. They are written with a view to:

(5) applying and possibly refining the judgement criteria in the specific context of the intervention, for instance:

- Do stakeholders spontaneously focus on the same judgement criteria as those selected for the evaluation? If not, why not?

(6) applying or developing the targets in the specific context of the intervention, for instance:

- Which are the areas / groups in the country with the best performance as regards the selected judgement criterion? Among them, which ones can legitimately be compared with targeted areas / groups?

8.2.2 Investigation areas

The evaluation team may consider collecting and analysing data at the level of the intervention as a whole, or investigating some areas more specifically, for instance:

- All sub-questions will be addressed through an investigation into a selection of countries and will include country notes and a visit to each selected country.

- In addition to using national statistics, the evaluation team will investigate a selection of districts respectively typical of (a) the targeted group /area, and (b) best performing groups / areas.
- The evaluation team will carefully select N projects which will be subject to an in-depth investigation in order to address some stages of the reasoning.

8.2.3 Information sources and tools

The design table identifies the sources of information to be used, such as:

- Statistics including context indicators available through international databases (see Volume 4)
- Management or monitoring databases
- Reports, reviews, audits, evaluations, articles, etc.
- Stakeholders' statements
- Experts' statements

In most instances a source of information is associated with an evaluation tool, for example a given category of stakeholder will be reached through interviews or focus groups, the opinion of experts will be obtained through email interaction or through a panel, etc.

8.3 Optimising the overall design

The overall evaluation method is finalised in a way which cuts across the evaluation questions and which makes a good enough mix of evaluation tools, considering the available time and resources.

Several iterations may be needed in order to allow the evaluation team to optimise the overall design whilst ensuring a high quality answer to each question.

Successive iterations require careful thought but are not time consuming and take place mainly during the inception and desk stages.

8.3.1 Combining tools and questions

The evaluation team draws up the list of all evaluation tools suggested in the design tables. Each tool is then considered from the viewpoint of its capacity to help answering several questions.

The evaluation team explains and justifies its main technical choices, with alternative options, pros and cons, and associated risks.

The evaluation team assesses whether the selected tools may reinforce each other, for instance:

- A series of interviews may help in identifying relevant texts which will be submitted to a documentary analysis
- A series of interviews may help in selecting the stakeholders to be invited to a focus group, and the issues to be discussed
- A series of interviews may help in refining the questions to be put to a sample of beneficiaries
- A series of case study monographs may be reviewed by a panel of experts with a view to strengthening the analysis and to deriving more accurate findings.

8.3.2 Preparing the overall assessment

The evaluation team examines all the design tables in a cross-cutting manner with a view to preparing its final synthesis, i.e. an overall assessment that draws upon the answers to all evaluation questions.

8.3.3 Allocating resources

In the process of designing its method, the evaluation team tries to adequately share its limited resources between questions. Some questions deserve to be addressed with costly tools such as a questionnaire survey of end users, a series of case studies, a series of focus groups, etc. Other questions should rather be answered on the basis of a documentary analysis only and a few interviews with EC and Government officials.

It is also wise to invest substantial resources in a question that raises feasibility problems.

A question is rarely unevaluable in the absolute. It is more likely to be an accumulation of difficulties and constraints that leads to feasibility problems. At the earliest stages of the evaluation, it is often possible to amend a difficult question so as to make it more evaluable. This can be done, for example, by limiting the scope of the question or choosing to apply it to a less distant effect or to likely effects if actual effects are not yet observable. Once a

difficult question has been validated, the evaluation team has to design an appropriate method, and to allocate adequate resources.

8.3.4 Cost and time constraints

Successive versions of the method are designed within the evaluation team until the following constraints are matched:

- The implementation of the tools fits into the overall time schedule of the evaluation process
- The cost of the tools (human resources, technical costs, travel and daily subsistence) fits into the overall budget of the evaluation
- The availability of qualified workforce in the field is sufficient for implementing the tools professionally.

8.4 Developing tools

The Volume 4 describes a series of tools that are frequently used with examples of use in the context of development aid evaluation.

Each tool implies a preparatory stage such as the writing of interview guides, the development and test of a questionnaire, etc. Tool development proceeds as follows:

- List of questions and steps of reasoning that have to be addressed by the tool
- Technical specifications for the implementation of the tool
- Expected risks that can compromise or undermine the implementation of the tool and the way to manage them
- Mode of reporting among the evaluation team
- Mode of reporting in the final report
- Responsibilities among the evaluation team members regarding the implementation of the tool
- Quality criteria and quality assessment process
- Time schedule
- Resources.

9 Data collection

SUMMARY

Data collection work plan

For the purpose of answering questions, the evaluation team collects data that are already available (secondary data) and applies data collection tools with a view to obtaining fresh information (primary data).

Frequent difficulties

Even if the data collection work plan has been carefully prepared, the evaluation team may encounter unexpected problems during its fieldwork. The difficulties encountered most frequently concern: access to informants, informants' unwillingness to cooperate, cultural distance, absence or weakness of information sought.

Reliability of collected data

Evaluation surveys involve several kinds of risk. The reliability of the data collected can be threatened by biases such as: bias of confirmation, bias of empathy, self-censorship, etc...

9.1 Data collection work plan

The data collection work plan finalises the methodological design (see Section 8.5).

Fresh (primary) data are collected by the means of tools such as: interviews, questionnaires, focus groups, field visits, direct observation, etc...

In order not to weary informants, and to avoid duplications and unnecessary costs, it is advisable to rely on existing (secondary) data as much as possible; in other words, on management and monitoring documents, studies and research dealing with the area under consideration, published statistical sources, previous evaluation reports and their annexes. Such information can be obtained at a low cost. It can help to provide partial answers to some of the questions asked.

If the evaluation team does not acknowledge relevant existing data, its work could be easily criticised by the stakeholders who consider those information as important. Then this would undermine the credibility of the whole evaluation.

Main channels for identifying and gathering secondary data

- Managers, implementing agencies, operators and partners
- Experts in the domain under consideration
- The Internet
- Statistical offices and monitoring bodies
- Scientific and professional literature.

9.2 Frequent difficulties and solutions

Even if the data collection programme has been wisely prepared, the evaluation team often encounters problems during its field work. The sections below present the most frequent difficulties and the way to solve them.

9.2.1 Access to informants

The sampling process proves to be difficult.

- Decide whether or not a reduced sample size is likely to provide statistically valid findings. If not, apply another technique such as the focus group.

An informant does not express him/herself freely.

- Focus interviews on facts rather than opinions
- Propose to keep collected information anonymous and explain how this will be secured.

An informant expresses him/herself in a way which seems purposely biased.

- Focus demands on facts, not on opinions.

9.2.2 Cultural gap

An informant or an information source can be accessed in the local language only.

- The evaluation team should include at least one member who is fluent in the local language (translation and interpretation always generate important information losses).

There is a large cultural gap between the evaluation team and the surveyed group.

- The evaluation team should include one or several members capable of bridging the gap between the two cultures.

9.2.3 Lack or weakness of data

An information source proves to be incomplete.

- If possible, extrapolate missing data and cross-check with other sources.

An information source proves to be unreliable.

- If possible, understand the biases, adjust data and cross-check with other sources.

Recommendations

Any evaluation creates a feeling of uncertainty, which makes some stakeholders reluctant to co-operate, if not hostile. In such cases keep a positive attitude, emphasise the intended use of the evaluation, promise impartiality, and focus on facts rather than opinions.

If an information source is not accessible or if a survey technique is not manageable, change the data collection work plan in order to collect similar information through other sources.

Pay sustained attention to biases and risks of unreliability. Strive to understand them. Report on them.

Avoid relying on one single information source in order to facilitate cross-checking at the analysis stage. This will also make it easier to manage if one of the sources cannot be used.

9.3 Reliability of data

While gathering information, the evaluation team faces various risks of biases which may undermine the reliability of collected data.

Why should biases be considered carefully?

- For improving the reliability of data collection
- For assessing the quality of the evaluation
- For understanding the limitations of conclusions which draw on unreliable data

Confirmation bias

This risk is a threat to all data collection approaches. It results from a tendency to seek out evidence that is consistent with the intervention logic, rather than evidence that could disprove it.

When subject to this bias, the evaluation team and informants tend to focus on intended effects and systematically to overlook external factors, unintended effects, negative effects, interactions with other policies, outside stakeholders, alternative implementation options, etc.

Self-censorship

In some instances, informants may be reluctant to freely answer questions, simply because they feel at risk. Then, they tend to rigidly express the views of their institution or their hierarchy.

This bias is combated by guaranteeing confidentiality and anonymity in the treatment of answers. The interviewer should also insist on factual questions and avoid collecting opinions.

Informants' strategy

Those who have stakes in the intervention may distort the information they provide, with the aim of obtaining evaluation conclusions closer to their views.

This bias will be reduced if the data collection work plan covers the whole range of stakeholders and if various sources of information are cross-checked.

Question induced answers

This bias and the following ones are frequent in interviews and questionnaires.

The way in which questions are asked by interviewers or the interviewer's reaction to answers can generate a bias which is either positive or negative.

Even the order of the questions in a questionnaire may change the substance of the answers.

This bias will be limited by having questionnaires designed and tested by experienced professionals.

Empathy bias

Especially in the case of interviews, the evaluation team has to create a friendly (empathetic) atmosphere, at least for the sake of achieving a high rate of answers and fast completion of the survey.

This introduces a systematic positive bias in the answers, which tends to overestimate the benefits of the intervention and to underestimate the role of external factors. Interviewees may not have a pre-determined opinion about the questions put to them. They try to make up their mind in a few seconds when responding to the interviewer or to the questionnaire. While doing so, they may be strongly influenced by the context.

Unrepresentative sample

This bias may be a matter of concern if the evaluation team generates quantitative data through a questionnaire survey. It should also be considered when using secondary data obtained from a questionnaire survey.

In this instance, the evaluation team should verify that the sample of surveyed informants is large enough and representative of the population as a whole.

Sample selection bias

People who agree to be interviewed may not be representative of the overall target audience.

This bias could be controlled by undertaking a special qualitative survey on a few "non-respondents", although this exercise brings additional costs.

Recommendations

Systematically mix positive and negative questions in order to reduce empathy bias and question induced bias.

Be highly credible when promising confidentiality and anonymity in order to limit respondents' self-censorship - and keep such promises strictly.

Never rely on a single category of stakeholder (e.g. programme managers, beneficiaries) in order to reduce strategic bias.

10 Analysis

SUMMARY

Strategy of analysis

Four strategies of analysis can be considered: change analysis, meta-analysis, attribution analysis, and contribution analysis. The last three allow answering cause-and-effect questions.

Analysis process

Once the strategy has been selected and the data collected, the analysis proceeds through all or part of the following four stages: data processing, exploration, explanation, confirmation.

Validity of analysis

Validity is achieved when conclusions and lessons are derived from findings in a way which ensures transferability (external validity), when findings are derived from data without any bias (internal validity), and when collected data reflect what is to be quantified or qualified without bias (construct validity).

10.1 Strategy of analysis

10.1.1 Change analysis

This approach informs on the change in quantified and/or qualified indicators over a given period of time. It does not strive to establish a cause-and-effect relationship between the observed change and the evaluated intervention.

Cause-and-effect questions

Cause-and-effect questions pertain to the effects of the evaluated intervention.

These questions call for an observation of change, and then an attribution of observed change to the intervention, or an analysis of the intervention's contribution to observed changes.

Effectiveness and impact questions tend to be cause-and-effect questions in the sense that they connect the evaluated intervention (the cause) and its effects. Efficiency and sustainability questions are also cause-and-effect questions since actual effects have to be analysed first, before being qualified as cost-effective or sustainable.

Generally speaking, relevance and coherence questions are not cause-and-effect questions.

10.1.2 Meta-analysis

This strategy extrapolates upon findings of other evaluations and studies, after having carefully checked their validity and transferability.

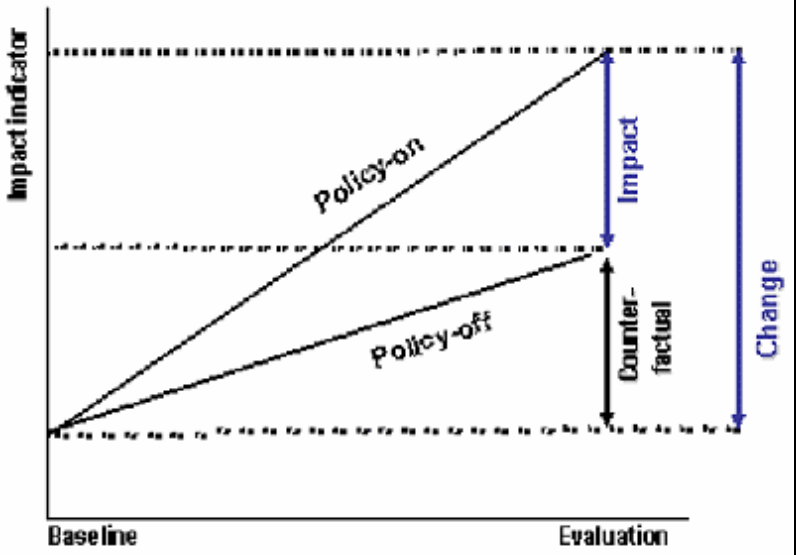
10.1.3 Attribution analysis

Attribution analysis aims to assess the proportion of observed change which can be attributed to the evaluated intervention. It involves building a counterfactual scenario (see Box 11).

The "policy-on" line shows the observed change, measured with an impact indicator, between the beginning of the evaluated period (baseline) and the date of the evaluation. For instance: local employment has increased; literacy has increased ... The impact accounts for only the share of this change that is attributable to the intervention.

The "policy-off" line, also called the counterfactual, is an estimate of what would have happened in the absence of the intervention. It can be obtained with appropriate approaches like a comparison group or modelling techniques. Impact is assessed by substracting the policy-off estimate from the observed change (policy-on).

Box 11 – Counterfactual scenario



Being derived from an estimate of the counterfactual, the impact is itself an estimate. In other words, impacts cannot be directly measured. They need to be derived from an analysis of impact indicators.

Only a counterfactual allows for a quantitative impact estimate. When successful, this approach has a high potential for learning and feedback. It is nevertheless relatively demanding in terms of data and skills, which makes it somewhat unusual in the evaluation practice in developing countries.

10.1.4 Contribution analysis

Contribution analysis aims to demonstrate whether or not the evaluated intervention is one of the causes of observed change. It relies upon chains of logical arguments that are verified through a careful confirmatory analysis.

It comprises the following successive steps:

- Refining the cause-and-effect chains which connect design and implementation on the one hand, and the evaluated

effect on the other. This task builds upon available explanations pertaining to the evaluated area. Explanations derive from the diagram of expected effects drawn in the first phase of the evaluation, from the evaluation team's expertise, and from exploratory analyses.

- Gathering evidence supporting each cause-and-effect assumption, including findings of similar studies, causal statements by interviewees, and evidence from in-depth inquiries.
- Gathering evidence supporting any alternative explanation (other interventions, external factors).
- Developing a step-by-step chain of arguments asserting that the intervention has (or has not) made a contribution, and possibly ranking the intervention among other contributions.
- Submitting the reasoning to systematic criticism until it is strong enough.

10.1.5 External factors

External factors are embedded in the context of the intervention and hinder or amplify the intended changes while being independent from the intervention itself. External factors are also called contextual, exogenous or confounding factors.

In a given evaluation, external factors are potentially numerous and it is crucial to highlight the most important ones.

Recommendations

Do not try to identify all possible external factors when clarifying the intervention logic in the structuring phase of the evaluation. They are simply too numerous. This task should be undertaken only when working on a given evaluation question.

10.2 Analysis process

It calls for interpretation and analysis to convert data into findings, which themselves call for a value judgement to be converted into conclusions. The analysis is undertaken on a question-by-question basis mainly, but some analyses may encompass several questions.

10.2.1 Data processing

Data are processed through operations such as cross-checking, comparison, clustering, listing, etc.

- Cross-checking is the use of several sources or types of data for establishing a fact.
- Comparison proceeds by building tables, graphs, maps and/or rankings. Data can be compared in one or several dimensions such as time, territories, administrative categories, socio-economic categories.
- Clustering proceeds by pooling data in accordance with predefined typologies.
- Listing proceeds by identifying the various dimensions of the needs, of the effects ...

Provisional findings emerge at this stage of the analysis. Further stages aim to deepen and to strengthen the findings.

10.2.2 Exploration

The exploratory analysis delves systematically into the collected data in order to discover new plausible explanations such as:

- New categories / typologies
- Unforeseen explanatory factors
- Factors favouring / constraining sustainability
- Unintended effects
- New cause-and-effect assumptions.

The analysis explores the set of data (quantitative and qualitative) with a view to identifying structures, differences, contrasts, similarities and correlations. For example, the analysis involves:

- Cross-cutting analyses of several case studies

- Statistical comparisons cutting across management data bases, statistical data bases, and/or the results of a questionnaire survey
- Comparisons between interviews and documents.

The approach is systematic and open-minded. Brainstorming techniques are appropriate.

10.2.3 Explanation

This next stage ensures that a sufficient understanding has been reached in terms of:

- Precisely defined concepts, categories and typologies
- Plausible cause-and-effect explanations
- Identification of key external factors and alternative explanations.

10.2.4 Confirmation

Provisional findings progressively emerge during the first phases of the evaluation team's work. They need to be confirmed by sound and credible controls. That is the role of the confirmatory analysis.

This analysis aims:

- To ensure that the findings are sound and able to withstand any criticism when the report is published
- To ensure that the findings are credible from the intended users' viewpoint
- In the particular case of cause-and-effect questions, to distinguish actual effects from observed change.

For a finding to be confirmed, it is systematically criticised by all possible means, e.g.

- If the finding derives from a statistical analysis, are the validity tests conclusive?
- If the finding was suggested by a case study, is it contradicted by another case study?
- If the finding derives from a survey, can it be explained by a bias in that survey?
- If the finding is based on an information source, is it contradicted by another source?

- Is the finding related to a change that can be explained by external factors that the evaluation team may have overlooked?
- Does the finding contradict expert opinions or lessons learned elsewhere and, if so, can this be explained?
- Do the members of the evaluation reference group have arguments to contradict the finding and, if so, are these arguments justified?

Strength of evidence

Findings should resist criticism, and therefore need to be supported by evidence, i.e. converging facts, records and/or statements. Four levels of strength can be considered:

- **Observed facts**

Factual evidence is the strongest.

Observed facts can be in the form of visit reports, photographs, management records or any kind of material trace.

- **Witness' statement**

Still very strong evidence, e.g. direct beneficiaries state that they have changed their attitude after participating in the programme.

- **Proxy**

This type of evidence is also called circumstantial evidence, e.g. during the past few months, several competitors of a subsidised firm collapsed, which indicates that the level of support was excessive and distorted competition.

The strength of this type of evidence depends upon the strength of the logical reasoning supporting the inference.

- **Reported statement**

An indirect statement is the weakest type of evidence, e.g. programme managers state that beneficiary enterprises have strongly improved their competitiveness.

The strength of this type of evidence depends upon the authoritativeness of the expert whose statement is used.

Recommendations

Devote relatively long interactions to the discussion of the final report in order to allow for a careful confirmatory analysis. Ensure that the evaluation team has put aside sufficient resources for that purpose.

Not all findings require the same level of confirmation. Concentrate efforts on findings that support the most controversial conclusions, the lessons that are the most likely to be transferred, or the recommendations that are the most difficult to accept.

In order to enhance the evaluation's credibility, it is valuable to list the criticisms that the findings withstood during the confirmatory analysis.

10.3 Validity of analysis

Validity is an essential quality of the analysis. Validity is achieved when:

- Conclusions and lessons are derived from findings in a way which ensures transferability (external validity)
- Findings are derived from data without any bias (internal validity)
- Collected data reflect the changes or needs that are to be evaluated without bias (construct validity).

10.3.1 External validity

External validity corresponds to the quality of an evaluation method which makes it possible to obtain findings that can be generalised to other groups, areas, periods, etc. External validity is fully achieved when the evaluation team can make it clear that a similar intervention implemented in another context would have the same effects under given conditions.

10.3.2 Internal validity

This is the quality of an evaluation method which, as far as possible, reduces the biases in data collection and analysis. Internal validity is fully achieved when the evaluation team

provides indisputable arguments showing that the findings derive from collected facts and statements.

10.3.3 Construct validity

This is the quality of an evaluation method which faithfully reflects the changes or needs that are to be evaluated. Construct validity is fully achieved when key concepts are clearly defined and when indicators reflect what they are meant to.

11 Judgement

SUMMARY

Conclusions and lessons

The conclusions provide clear answers to the questions asked at the beginning of the evaluation. They involve value judgements on the merits and worth of the intervention.

The lessons are conclusions that can be transferred to the next cycles of the same intervention or to other interventions.

Recommendations

The recommendations do not involve a value judgement strictly speaking, but they derive from conclusions. They aim at improving or reforming the evaluated intervention, or at preparing the design of a new intervention for the next cycle.

11.1 Conclusions and lessons

The conclusions are based on judgement criteria

As far as possible the evaluation report distinguishes the findings (which follow only from facts and analysis) and the conclusions (which entail a value judgement).

To formulate its conclusions, the evaluation team applies the judgement criteria (also called "reasoned assessment criteria") that were agreed upon in the first phase (desk) of the evaluation (see Section 7.1). Data collection and analysis are structured according to these criteria. As long as this is possible, the findings are compared against targets.

This approach necessarily makes the judgement criteria explicit, which in turn reinforces the quality of the evaluation. Recognising that the conclusions include a value judgment also conducts to better respect ethical principles presented hereafter.

Ethical principles

- **Responsibility**

In the synthesis phase, the evaluation team applies the judgement criteria agreed on, as faithfully as possible, and produces its own conclusions. The conclusions are discussed within the reference group but remain the entire responsibility of the evaluation team.

- **Legitimacy**

The questions and judgement criteria take into account the needs and the standpoint of the public institution that initiated the evaluation

The members of the reference group contribute different points of view, which reinforces the legitimacy of the evaluation

The evaluation team holds interviews, which may enable it to identify other points that were not expressed by the reference group members. It makes them known in reference group meetings and may take them into account.

- **Impartiality**

The players ensure that all the opinions are heard, even if they are not conveyed by the loudest voices or the majority. They are aware of asymmetrical power relations, and they correct the biases arising from such imbalances.

Evaluation team members are familiar with and respectful of beliefs, manners and customs of concerned groups.

Dissenting views are mentioned in the report and at reference group meetings.

- **Protection of individuals**

The conclusions concern the merits of the evaluated intervention, not the persons who implement it or benefit from it.

At the stage of the draft final report, the evaluation team may have to refine its judgement criteria and targets. In such a case, the changes are specified and discussed with the reference group.

Answering the questions and drawing general conclusions

Conclusions provide clear answers to the questions asked at the beginning of the evaluation.

One chapter of the report organises the conclusions in clusters in order to provide an overall assessment of the evaluated intervention. This chapter does not follow the order of the questions or that of the evaluation criteria (effectiveness, efficiency, coherence, etc.).

Presentation of the conclusions

The report contains a list of conclusions referring to the sections of the report or to annexes showing how the conclusions derive from data, interpretations, analysis and judgement criteria. The report includes a self-assessment of the methodological limitations that may restrain the range or use of certain conclusions.

The report identifies the lessons learnt, i.e. the conclusions that can be transferred to the next cycles of the same intervention or to other interventions.

A paragraph or sub-chapter picks up the 3 or 4 major conclusions organised by order of importance, while avoiding being repetitive. This practice allows to better communicate the evaluation messages that are addressed to policy makers within the Commission.

Warning!

- The evaluation team and those in charge of quality assurance are advised to carefully reread the final report and to eliminate any unessential and/or unintended value judgements.
- The evaluation team presents its conclusions in a balanced way, without systematically favouring the negative or the positive conclusions.

11.2 Recommendations

11.2.1 How to draft and present them?

The recommendations must be related to the conclusions without replicating them. A recommendation derives directly from one or more conclusions.

> **Warning!**
>
> If a recommendation does not clearly derive from the conclusions, it probably reflects preconceived ideas or the tactical interests of one of the stakeholders. Its presence in the report could then discredit the evaluation.

The recommendations are clustered and prioritised. The report mentions the addressees of the recommendations, e.g. EC Delegation, services in charge of designing the next intervention, etc.

The recommendations are useful, operational and feasible, and the conditions of implementation are specified. Wherever possible and relevant, the main recommendations are presented in the form of options with the conditions related to each option, as well as the predictable consequences of the implementation of each option. The recommendations are presented in a specific chapter. This chapter highlights the recommendations derived from the three or four main conclusions.

11.2.2 How to promote them?

The recommendations are valuable as far as they are considered and, if possible, owned by their addressees. To promote their take-up, the manager drafts a *fiche contradictoire* in order to:

- List the recommendations in a shortened form
- Collect the addressees' responses
- Inform on actual follow-up to the recommendations, if any.

12 Quality assurance

SUMMARY

Quality assurance consists of establishing quality check-points at the main phases of the process, defining the responsibilities of quality assurance, and establishing the rules of the game in case of quality defects.

Quality assurance enhances the conclusions of the evaluation in users' eyes by showing that they arise from an impartial and rigorous work that meets professional standards.

Quality is constructed gradually so as to avoid discovering a major quality defect in the final report stage.

The quality assurance rules clarify the interactions between the evaluation manager, the reference group and the evaluation team.

12.1 Rules of the game

The quality assurance rules are specified in the terms of reference and pointed out when the evaluation team is engaged.

Successive quality items are secured at each step in order to go on solid bases for the next step.

These rules concern:

- Approval of documents
- Quality criteria
- Dissemination of the quality assessment of the final report.

The quality assurance process benefits from the contribution of all actors whilst limiting the potential conflicts that might arise between them, e.g.

- The actors who hold responsibilities in the evaluated intervention (the evaluees) are in a good position for assessing the relevance of data collected and the fairness of interpretations. Their assessment may, however, be distorted by a confirmation bias. They should therefore be given only an advisory role, e.g. through their participation in the reference group.

- The evaluation team leader has a major role in assuring quality, especially in designing an adequate method, and in securing the accuracy and fairness of the report. He may, however, overweight the views of powerful stakeholders, especially those who are likely to commission other evaluations in the future. The evaluation manager's quality assessment helps to prevent such a risk.

Quality criteria

The quality of the evaluation report is rated along with nine criteria by means of a grid which involves five levels: Excellent, Very good, Fair, Poor and Unacceptable.

Criteria 1: Meeting needs

Does the report adequately address the information needs of the commissioning body and fit the terms of reference?

Criteria 2: Relevant scope

Are the rationale of the intervention and its set of outputs, results and impacts examined fully, including both intended and unexpected policy interactions and consequences?

Criteria 3: Defensible design

Is the evaluation design appropriate and adequate to ensure that the full set of findings, along with methodological limitations, is made accessible for answering the main evaluation questions?

Criteria 4: Reliable data

Are the primary and secondary data selected adequate? Are they sufficiently reliable for their intended use?

Criteria 5: Sound analysis

Is the analysis of quantitative and qualitative information appropriately and systematically analysed according to the state of the art so that evaluation questions are answered in a valid way?

Criteria 6: Credible findings

Do findings follow logically from, and are justified by the data analysis and interpretations based on carefully described assumptions and rationale?

Criteria 7: Valid conclusions

Does the report provide clear conclusions? Are conclusions based on credible findings?

Criteria 8: Useful recommendations

Are recommendations fair, unbiased by personal or stakeholders' views, and sufficiently detailed to be operationally applicable?

Criteria 9: Clear report

Does the report clearly describe the intervention being evaluated, including its context and purpose, together with the process and findings of the evaluation, so that information provided can easily be understood?

A simplified grid is presented in Volume 3 for the evaluation of projects and programmes.

Criteria 1 and 2 are assessed at the stage of the inception report.

The desk report is quality assessed for criteria 1 to 3, and for criteria 4 as regards data which have already been collected.

All 1 to 9 criteria are checked at the stage of the final report (also at the stage of the draft final report).

On the website: suggestions for assessing each of the 9 criteria

12.2 Approval of deliveries

The approval of a report is a formal recognition by the evaluation manager that the delivered document has an adequate content and meets the applicable quality criteria in methodological terms

12.2.1 What are the approval steps?

Generally, a document is approved in the following way:

- The manager checks that the document has the required form and content, and has no major quality defect. If it does, he immediately requests a new version.

- The manager asks for the opinion of the individuals or services that have to be consulted, either at a meeting (reference group) or by email. He gives a deadline after which there is no guarantee that comments will be taken on board.
- The manager lists his/her own requirements regarding quality improvements, attaches other requests received, distinguishes requests applying to the methodological quality from requests applying to the substance of the text, and sends all the comments to the evaluation team.
- The evaluation team takes into account all the requests for quality improvements, either by amending its document accordingly or by mentioning in an annex how the request will be treated at a later stage, or else by explaining in an annex why the request could not be satisfied.
- A new version of the document is soon presented to the manager who accepts or refuses it.
- When approval is required, it is also a prerequisite for moving on to the next stage.

12.2.2 Which documents are approved and by whom?

The following documents require approval:

- Inception report
- First phase report (desk)
- Draft final report
- Final report.

Approval is generally performed by the evaluation manager and may be confirmed by his/her superior.

Glossary

The vocabulary and the terminology developed in the four guides are defined by OECD on the one hand and the European Commission on the other hand.

OECD definitions

In 2002 the OECD published a glossary of key terms in evaluation and results-based management (Development Aid Committee - Working Party on Aid Evaluation). Initially issued in English, French and Spanish, it has been subsequently translated in many other languages. It is part of a common effort of the development aid community, including the European Community, to reduce the terminological confusion and to promote common vocabulary.

EC definitions

The Financial Regulation establishes a Boxwork which applies to all EC policies. This Boxwork has been translated into evaluation guidelines and a glossary issued by the Directorate General Budget.

Comparing definitions

Both OECD and EC definitions should be regarded as equally legitimate in the context of EC external assistance evaluation. An in-depth comparison has been undertaken with the result that: ***almost all terms are defined consistently.***

Only four couples of definitions are different enough to raise significant risks of misunderstanding:

Risks of misunderstanding (four terms)		
	OECD	**EC**
Impact	… long term effects produced by a development intervention…	A general term used to describe the effects of an intervention on society …
Outcome	The likely or achieved short-term and medium-term effects of an intervention's outputs.	The longer-term impact usually expressed in terms of broad socio-economic consequences
Result	…The output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention.	The initial impact of an intervention …
Efficiency	A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results (considering OECD's definition of results, efficiency may relate to outputs or to any level of effect)	The extent to which the desired effects are achieved at a reasonable cost (a definition which does not cover outputs)

The following options have been taken in order to design these guidelines:

Options taken		
	Basic glossary	**Definition**
Impact	OECD	… long term effects produced by a development intervention …
Outcome	The term has been neither used nor defined	
Result	EC	The initial impact of an intervention …
Efficiency	Merged (based on OECD)	The extent to which outputs and/or the desired effects are achieved with the lowest possible use of resources/inputs (funds, expertise, time, administrative costs, etc.)

In other publications issued by The European Commission, the term results can be used in the wider sense, as defined by OECD-DAC, e.g. Result-Oriented Monitoring.

Available on Europa website

http://ec.europa.eu/europeaid/index_en.htm

Numerous examples and check-lists, especially:

Examples of questions for a country level evaluation, a regional evaluation, a global, thematic or sectoral evaluation

Examples of articulation between questions, judgement criteria and indicators

Examples of design table and developing a tool

Quality assessment grid and suggestion for filling it

Examples of summaries and articles about an evaluation.

List of Boxes

EUROPEAN COMMISSION

Evaluation Methods for the European Union's External Assistance

2006 – 97 pp. – 21 x 14,8 cm

Luxemburg: Office for Official Publications of the European Communities.

ISBN: 92-79-00681-9