Sida

Looking Back, Moving Forward

Sida Evaluation Manual

2nd revised edition

Looking Back, Moving Forward

Sida Evaluation Manual

2nd revised edition

Stefan Molund
Göran Schill

Part One:
Concepts and Issues

Chapter 1: What is Evaluation?

Boxes

Chapter 2: Evaluation Criteria

Chapter 3: Evaluating Poverty Reduction

Part Two:
The Evaluation Process Step by Step

Useful Evaluation Sites on the Internet

www.oecd.org Website of the OECD/DAC Network on Development Evaluation. Contains useful downloads as well as links to the websites of the evaluation departments of multilateral and bilateral development organisations.

www.worldbank.org Website of the World Bank with its Independent Evaluation Group (IEG). Contains a wealth of useful publications.

www.eldis.org Gateway for information on development issues hosted by the Institute of Developments Studies (IDS), University of Sussex. Supported by Sida and other development agencies. A well stocked database with many useful online publications relevant to evaluators.

www.alnap.org The Active Learning Network for Accountability and Performance in Humanitarian Action – an international interagency forum working to improve learning, accountability and quality across the humanitarian sector. Features a standard evaluation manual for humanitarian assistance.

www.policy-evaluation.org WWW Virtual Library: Evaluation. Worldwide gateway for evaluation information with links to evaluation societies, national courts of audit, and evaluation departments in international organisations and the EU. Plenty of educational materials for downloading.

www.mande.co.uk MandE News. News service focusing on the development of methods for monitoring and evaluation in development projects and programmes. Much information about publications and events.

Instructions for Use

This is a manual for evaluation of development interventions. It is intended primarily for Sida staff, but may also be useful to Sida's co-operation partners and independent evaluators engaged to evaluate activities supported by Sida.

It consists of two main parts. The first deals with the concept of evaluation, roles and relationships in evaluation, and the evaluation criteria and standards of performance employed in development co-operation. The second is a step-by-step guide for Sida programme officers and others involved in the management of evaluations initiated by Sida or its partners.

Although it can be read in its entirety all at once, most readers will probably use it in a piecemeal fashion. A reader who wants a rapid overview of basic concepts in the evaluation of development interventions should consult part one. A reader who is engaged in the management of an evaluation, may turn directly to part two and return to part one as need arises.

Some familiarity with basic development co-operation terminology is helpful, but not essential. What is required, on the other hand, is a readiness to engage with new concepts, some of which may seem complicated at first glance. A certain tolerance of ambiguity is also useful. Although evaluation is a professional field of its own, many of its key terms do not have standard definitions.

The manual adheres to the terminological conventions recommended by the OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management, which is included as an annex. Yet, in some cases it also takes notice of alternative usage. For Sida evaluation managers who must agree on conceptual matters with co-operation partners and professional evaluators some knowledge of the variability of international evaluation terminology can be very useful.

Two limitations should be noted. One is that this is a handbook for evaluation managers rather than evaluators. While it deals extensively with conceptual and organisational matters, it has little to say about the technicalities of the evaluation research process. Still, evaluators may find it useful as a reference point for discussions about evaluation with Sida and its partners.

A second limitation is that it focuses on matters that are common to all or most kinds of development evaluations. It does not deal with the peculiarities of the different sub-fields of development evaluation. Readers seeking information about matters that are unique to the evaluation of humanitarian assistance, research co-operation, programme support, and so on, must turn to other manuals.

Part One:
Concepts and Issues

Chapter 1

This chapter explains the concept of evaluation and deals with some basic issues concerning evaluation in development co-operation. More specifically, it answers the following questions:

- What is evaluation?
- What is monitoring and how does it differ from evaluation?
- What are the purposes and uses of evaluation?
- What is an external evaluation?
- What is an internal evaluation?
- What is a participatory evaluation?
- How can stakeholders participate in evaluations?
- How does evaluation fit into a framework of development co-operation partnership?
- What is a good evaluation?

What is Evaluation?

The definition in Sida's Evaluation Policy[3] is much the same:

"…an evaluation is a careful and systematic retrospective assessment of the design, implementation, and results of development activities."

These definitions differ from the broader one in two important respects:

- An evaluation is an assessment of ongoing or completed activities, not activities that are still at the planning stage. Project appraisals, feasibility studies, and other assessments of proposed future activities are sometimes called ex ante evaluations or prospective evaluations. Without further specification, however, the term evaluation refers solely to retrospectively based assessments of ongoing or completed activities.

- No assessment is an evaluation unless it satisfies certain quality standards. To be considered an evaluation, an assessment must be carried out systematically and with due concern for factual accuracy and impartiality.

1.1 The concept of evaluation

The word evaluation has many meanings. In the widest sense evaluation is:

"…the process of determining the merit, worth, or value of something…"[1]

So defined, evaluation covers a wide range of activities, many of which can also be described by other words, such as appraise, assess, examine, judge, rate, review, and test. In development co-operation, however, the word evaluation is understood more narrowly. According to the standard definition of the Development Assistance Committee (DAC) of the OECD an evaluation is:

"… a systematic and objective assessment of an ongoing or completed project, program or policy, its design, implementation and results."[2]

Evaluation, as we use the term, is first and foremost a reality test, a learning mechanism that provides feedback on the results of action in relation to prior objectives, plans, expectations or standards of performance. The requirement

[1] Scriven, M. Evaluation Thesaurus (4th ed.). Newbury Park, CA: Sage, 1991, p. 139.
[2] OECD/DAC. Development Assistance Manual. Paris, 2002.
[3] www.sida.se/publications

that evaluations should be systematic and objective derives directly from the function of evaluation as a reality test. To serve as a probe on the realism and accuracy of plans and expectations, an evaluation must use sound and transparent methods of observation and analysis.

The term evaluation is sometimes used exclusively for assessments made at the very end of an activity or later. Common in project cycle management, this understanding of the term reflects the idea of an evaluation as a comprehensive stocktaking when an activity has run its entire course. In this manual, however, the term covers both assessments of ongoing and completed activities.

An evaluation of activities in progress is referred to as an *interim evaluation*. An evaluation that is carried out when an activity is completed is an *end-of-project* or *end-of-programme evaluation*. An evaluation at some later point in time is known as an *ex-post evaluation*. The term *review* is often used to describe an assessment that is less ambitious in scope or depth than an evaluation.

With regard to focus, there are two broad types of evaluation, *process evaluation* and *impact evaluation*. A process evaluation deals with the planning and implementation of an activity as well as with outputs and other intermediary results. An impact evaluation, by contrast, is mainly concerned with the effects – outcomes and impacts – brought about through the use of outputs. Whereas interim evaluations tend to be mainly process evaluations, end-of-programme evaluations and ex post evaluations are more likely to focus on effects.

The object of evaluation – what textbooks refer to as the evaluand – provides a further ground for classification. In development co-operation we talk about project evaluations, programme evaluations, sector evaluations, country evaluations, and so forth. Definitions can be found in the glossary. Development intervention evaluation is a blanket term for all of them.

1.2 Monitoring and evaluation

Evaluation should be distinguished from monitoring, the continuous follow-up of activities and results in relation to pre-set targets and objectives. Monitoring is sometimes regarded as a type or component of evaluation, but in development co-operation we usually regard it as a separate activity. The distinction is primarily one of analytical depth. Whereas monitoring may be nothing more than a simple recording of activities and results against plans and budgets, evaluation probes deeper.

"Monitoring and evaluation are interdependent"

Although monitoring signals failures to reach targets and other problems to be tackled along the way, it can usually not explain why a particular problem has arisen, or why a particular outcome has occurred or failed to occur. To deal with such questions of cause and effect, an evaluation is normally required. An evaluation may also help us gain a better understanding of how a development intervention relates to its social and cultural environment, or it can be used to examine the relevance of a particular intervention to broader development concerns. Furthermore,

Box 1

Monitoring and evaluation

MONITORING	EVALUATION
Continuous or periodic	Episodic, ad hoc
Programme objectives taken as given	Programme objectives assessed in relation to higher-level goals or to the development problem to be solved
Pre-defined indicators of progress assumed to be appropriate	Validity and relevance of pre-defined indicators open to question
Tracks progress against small number of pre-defined indicators	Deals with wide range of issues
Focus on intended results	Identifies both unintended and intended results.
Quantitative methods	Qualitative and quantitative methods
Data routinely collected	Multiple sources of data
Does not answer causal questions	Provides answers to causal questions
Usually an internal management function	Often done by external evaluators and often initiated by external agents

while the objectives of a development activity are taken for granted in monitoring, and progress indicators are assumed to be valid and relevant, evaluations may question both. Box 1 presents a more extensive list of contrasts.

Sida regards evaluation as a complement to monitoring that should be used selectively to deal with problems that monitoring cannot adequately handle[4]. For an evaluation to be feasible, however, monitoring data may be necessary. If an intervention has not been properly monitored from start, it may not be possible to subsequently evaluate satisfactorily. Just as monitoring needs evaluation as its complement, evaluation requires support from monitoring.

As shown in Box 2, an intervention can be monitored at different levels. The monitoring of inputs and outputs, input-output monitoring for short, keeps track of the transformation of financial resources and other inputs into goods and

services. The monitoring of outcomes and impacts – outcomes more often than impacts – seeks to register the intended effects of delivered goods and services on targeted groups or systems. Outcome monitoring, sometimes referred to as beneficiary monitoring, measures the extent to which intended beneficiaries have access to outputs and are able to put them to good use.

In development co-operation, more effort is usually devoted to input-output monitoring than to the monitoring of outcomes and impacts. But both types of monitoring are important. Without carefully tracking the processes leading from inputs to outputs, a programme may fail to reach its production targets. Without outcome-impact monitoring, it runs the risk of producing outputs that have little or no developmental effect. For the purpose of evaluation, information about outcomes and impacts is often crucially important.

---

[4] Sida at Work – A Guide to Principles, Procedures, and Working Methods. Stockholm, 2003.

Levels of monitoring

IMPACTS	Effects on life chances and living standards (inter alia) – infant child mortality – prevalence of a specific disease
OUTCOMES	Access, usage, and satisfaction of users – number of children vaccinated – percentage of vaccinated children within 5 km of health centre
OUTPUTS	Goods and services produced – number of nurses – availability of medicine
INPUTS	Financial and other resources – spending in primary health care

Source: Africa Forum on Poverty Reduction Strategy. June 5–8, 2000, Côte d'Ivoire.

1.3 Accountability and learning

In section 1.1, evaluation is described in general terms as a research tool providing feedback on the results of action with regard to prior objectives, plans, expectations or standards of performance. However, nothing further is said about the practical purpose for which it can be employed.

In development co-operation, the standard answer to the question of the practical purpose of evaluation is that it serves two broad types of ends: accountability and learning. As briefly discussed in section 1.4 below, evaluation can also be used for other types of purpose, but these two are usually taken to be the main ones.

Accountability is a relationship which exists where one party – the principal – has delegated tasks to a second party – the agent – and the latter is required to report back to the former about the implementation and results of those tasks. As commonly understood, accountability consists of two separate components. One is the agent's answerability vis-à-vis the principal, the other the power of the principal to initiate remedial action or impose sanctions in case the agent fails to carry out his obligations as they have been agreed.

The content of the reporting varies from case to case. An important distinction is that between *financial accountability*, which is answerability for the allocation, disbursement and utilisation of funds, and *performance accountability,* which concerns results. Evaluation, in so far as it serves as a tool for accountability, provides information for reporting about performance and results. It is less concerned with financial accountability, which is mainly the province of auditors and accountants. As explained in section 1.2, evaluation is particularly useful when results must be analysed in depth. For routine reporting of outputs and easily measured outcomes, monitoring may serve equally well.

In general terms, what an evaluation for accountability seeks to find out is whether the organisations that are responsible for the evaluated intervention have done as good a job as possible under the circumstances. This means trying to find out if and to what extent the intervention has achieved the results that it was intended to achieve or that it could reasonably have been expected to achieve. It is also likely to involve an assessment of the quality of the processes of planning and implementation. When results are difficult to measure – a common situation in development co-operation – an evaluation initiated for purposes of accountability

may focus entirely on questions about process. Box 3 lists some of the issues that a process-oriented accountability assessment might consider.

When the purpose of evaluation is learning, on the other hand, the study is expected to produce substantive ideas on how to improve the reviewed activity or similar activities. Although learning, in itself, may be regarded as valuable, its real importance lies in the translation of new knowledge into better practice. If we did not expect learning to enhance organisational performance or contribute to practical improvement in some other way, we would hardly see it as a justification for spending money and effort on evaluations. Evaluations that are primarily meant to contribute to learning are often called *formative* evaluations, whereas evaluations for accountability are described as *summative* evaluations.

Note that the distinction between accountability and learning, summative and formative, refers to the purpose and use of evaluations rather than to their contents. An evaluation that is meant to satisfy a requirement for accountability may of course raise very different questions than an evaluation intended for learning. Still, there are many evaluation questions that are likely to be important for both purposes. For example, most of the questions in Box 3 could appear in an evaluation intended for learning as well as in an accountability evaluation. Indeed, in many cases different audiences may use one and the same evaluation for different practical ends. It is not unusual that an evaluation, used by those who are responsible for the evaluated activity for improvement and learning, serves a purpose of accountability in relation to principals and the general public. As evaluator Robert Stake is reported to have said[5]:

> "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative"

1.4 Use and abuse

Note also that evaluations can be used for purposes other than those of accountability and learning. Box 4 contains a classification of actual uses of evaluations adapted from a standard textbook. While the first two of

[5] Scriven, M. Evaluation Thesaurus (4th ed.). Newbury Park, CA: Sage, 1991, p. 169.

the uses identified under this scheme – the instrumental and the conceptual – are readily subsumed under accountability and learning, the remaining ones do not fit these concepts equally well. The ritual use is clearly the most divergent, but using an evaluation to seek support for a decision that has already been made on other grounds, as in the case of legitimisation, or using it for tactical ends, could also be inconsistent with both accountability and learning.

The box may serve as a reminder that not every reason for conducting an evaluation is equally respectable. Evaluation is based on principles of impartiality, transparency, and open discussion, but these values can be subverted by other concerns. As evaluations may have far-reaching consequences for stakeholders' reputations and resources, this is entirely understandable. Still, it is important to keep in mind that there is a politics of evaluation, and that evaluation is often something else or something more than just a useful tool for social engineering.

We should also consider the possibility that evaluations may be initiated for purely formal reasons or because it is vaguely felt that they ought to be done. Although evaluations carried out on such non-rational grounds may be rare, it is not uncommon that evaluations are initiated

Varieties of use

INSTRUMENTAL USE
The findings and recommendations of the evaluation are directly used as an input to decision-making concerning the evaluated intervention or some related intervention or set of interventions. The decision may follow the conclusions or recommendations of the evaluation, but even if it does not, the evaluation is seriously taken into account in the deliberations leading to the decision.

CONCEPTUAL USE
The use of evaluation to obtain a deeper understanding of an activity or type of activity. It is expected that the evaluation will feed into ongoing processes of organisational learning and eventually have a useful impact on practice, but the evaluation itself is not directly used in decision-making.

LEGITIMISATION
The evaluation is used to mobilise authoritative support for views that are held regardless of the evaluation. It is intended to justify a particular interest, policy, or point of view, rather than to find answers to unresolved questions or provide solutions to outstanding problems.

TACTICAL USE
The evaluation is used to gain time, evade responsibility, or, perhaps, create an opportunity for dialogue among stakeholders. It is intended to convince users that matters are under control; that the programme is responsibly administered; and so on. The impact of the evaluation, if any, derives from the fact that it is carried out, rather than from its findings and conclusions.

RITUAL USE
There are two kinds of ritual use. In the one case, where ritual is understood as nothing more than prescribed formal behaviour – 'empty ritual', as it were – evaluations are carried out for no other reason than that they should be done. In the other case, where rituals are understood more anthropologically as symbolic expressions of collective beliefs and values, a ritual use of evaluation is one whereby the participants remind themselves of the larger meaning of the evaluated activities. The evaluation gives them an opportunity to focus on the 'big picture'.

PROCESS USE
Ritual use and tactical use can be seen as special cases of a wider category of use called process use. When we talk about process use, we refer to the use of the evaluation process itself rather than its products. Evaluation processes can be used to create shared understanding or boost confidence and morale. Although evaluations may be designed with such purposes in view, process use is often an unintended by-product of an evaluation serving another purpose.

Adapted from Evert Vedung. Public Policy and Program Evaluation. New Brunswick: Transaction Publishers, 1997.

routinely and without a sufficiently considered purpose. When this is the case the information produced is not likely to be useful to anyone, and the evaluation process itself is a waste of time and resources. In this manual we stress the point that clearly and transparently defining the practical purpose of the study – its intended use by identified actors – is the first and most important of all the steps in a process of planning and conducting an evaluation.

For a development co-operation agency such as Sida clearly articulating the purpose of the evaluations that it initiates is always advisable. What Sida's co-operation partners and other stakeholders will be particularly concerned about, is not if Sida's purpose in carrying out an evaluation is accountability, learning or something else, but, more precisely, how the evaluation is intended to be used, and how this use is likely to affect them. Transparency on this point is essential. For development agencies the best way of promoting transparency is to invite partners and other key stakeholders to participate in the making of the evaluation.

1.5 External and internal evaluation

We commonly make a distinction between external and internal evaluation. As defined by the OECD/DAC evaluation glossary included in this manual, an evaluation is external when it is conducted by entities and/or individuals outside the organisations responsible for the evaluated activities. In development co-operation an evaluation is external when those doing it stand outside the partner country organisation in charge of the evaluated activities, the supporting donor organisations, as well as any implementing organisation, such as a firm of consultants or an implementing NGO.

An internal evaluation, by contrast, is an evaluation conducted by entities or individuals reporting to the management of one or several of the organisations involved in the evaluated activities. As internal evaluations are conducted by units that are constituent parts of those organisations, they are often also described as self-evaluations.

The distinction between external and internal evaluation is closely associated with notions of independence, impartiality and bias. As we usually understand the term, an evaluation is not just any assessment of the merits of an activity, but one that aims to be as objective and impartial as possible. The requirement for objectivity and impartiality is always strong where evaluations serve a purpose of accountability, and it can be equally strong when learning and the creation of new knowledge is the purpose.

The link between the organisational location of an evaluation, on the one hand, and its objectivity and impartiality, on the other, is not difficult to understand. We have all experienced situations where perceptions and judgements were distorted by partial perspectives or interests. When assessing the credibility of a statement we routinely assess the likelihood that the person making it is biased by self-interest, loyalties, or other disorting factors. The requirement that evaluations should be independent of the evaluated activities is an instance of the same logic.

Yet, independence and freedom from bias is just one of several quality requirements in evaluation. What we want from an evaluation are accurate and relevant answers to the questions that the evaluation raises. Freedom from bias is not a guarantee of such answers. To be able to answer the evaluation questions accurately the evaluator must also have access to relevant information and be well-informed with regard to the activities or type of activities to be assessed.

Independence cannot compensate for lack of knowledge and information.

This is where internal evaluation comes into the picture. While the external evaluator tends to have an advantage over the internal evaluator with regard to objectivity and bias, and is usually regarded as the more credible of the two, the internal evaluator is sometimes better placed to understand the workings of the organisations and activities to be assessed. The internal evaluator may also be in a better position to communicate evaluation findings to the intended users of the evaluators.

Note, however, that the distinction between external and internal evaluation is a simplification of what is usually a more complex situation. In the Swedish system for evaluating development co-operation there are at least three levels of independence. At the one extreme there is the Swedish Agency for Development Evaluation (SADEV), a separate government agency devoted entirely to evaluation. This agency has its own board and budget and is separate from all organisations responsible for the implementation of Swedish development assistance. At the other extreme, there is a system of evaluation where Sida's line departments hire external evaluators to assess the progress of projects and programs in their respective portfolios. In between these two, we have Sida's own evaluation department, the Department for Evaluation and Internal Audit (UTV), focusing on larger issues of strategic importance to Sida as a whole. Reporting directly to Sida's Board, UTV it is neither internal nor external, but a little of both.

The Swedish situation is not unique. Although separate entities like SADEV are uncommon, a division between a central independent or semi-independent evaluation office and a system of self-evaluation by operative departments is not unusual. Similar divisions can also be found in many multilateral organisations.

Note also the distinction between independence in deciding what to evaluate, on the one hand, and independence in the design and implementation of the evaluation research process, on the other. Discussions of independence in evaluation often focus on the second aspect, sometimes completely ignoring the first. Yet, both are obviously important. From an accountability point of view, an evaluation where the prerogative of asking the questions rests with external actors should clearly carry more weight than one where this task is vested with the organisations responsible for the evaluated activities. In many cases, the perspectives and interests influencing the formulation of the evaluation questions is a more important source of bias than the methods used in answering them.

> "The perspectives influencing the questions can be a more important source of bias than the methods used in answering them"

A more elaborate scheme for assessing the independence of evaluation units is provided the World Bank[6]:

- **ORGANIZATIONAL INDEPENDENCE.** Direct line of reporting between evaluation unit and the management or board of the institution. Evaluation unit located outside operational staff and line management function. Evaluation unit removed from political pressure and able to operate without fear of repercussions. Recruitment and promotion of staff based on merit.

- **BEHAVIOURAL INDEPENDENCE.** Ability and willingness of the evaluation unit to produce strong and uncompromising reports. Transparency in reporting.

- **AVOIDANCE OF CONFLICTS OF INTEREST.** Extent to which stakeholders are consulted in evaluations to ensure against evaluator bias. Extent to which there are procedures to prevent staff from getting involved in evaluations of activities where they or their

---

[6] This is a slightly abbreviated version of a scheme outlined in OED Reach, February 24, 2003, World Bank.

close associates have, have had, or in the future may expect to have a substantial interest.

■ **PROTECTION FROM EXTERNAL INFLUENCE**. Extent to which the evaluation unit is able to decide on the design, scope, timing and conduct of evaluations without undue interference. Extent to which the evaluation unit has control over staff hiring, firing and promotion. Extent to which the employment of evaluation staff is not jeopardized for reasons other than competency and need for evaluation services. Extent to which the evaluation unit can fulfil its obligations unimpeded by restrictions on funds. Extent to which the judgements of evaluators regarding the contents of reports is protected from overruling by external authorities.

Although designed for the assessment of evaluation units inside donor organisations this scheme is also relevant for the assessment of free-standing units like the Swedish SADEV. Obviously, it is relevant for the assessment of evaluation units in developing countries as well as for donor evaluation units.

1.6 Participatory evaluation

To make our inventory of types of evaluation by independence complete, we need to discuss still another type of evaluation known as participatory evaluation. While both external and internal evaluation can be described as forms of expert evaluation, participatory evaluation is an evaluation modality where the distinction between expert and layperson, researcher and researched, is de-emphasised or, at the limit, abolished altogether. In the one case the evaluators are expected to make an independent expert assessment of the merits of the

evaluated activities, in the other, their role is mainly to assist program participants in making their own assessment. A more detailed overview of the differences between expert evaluations and participatory evaluations is provided in Box 7.

The case for participation in evaluation is much the same as the case for participation in development co-operation generally. Participation means putting ordinary people first, and redefining the roles of experts and laypersons. Participation can be an end in itself, an expression of the right for people to have a voice in matters that significantly affect them. It can also be justified in instrumental terms, in which case it helps mobilise local knowledge and makes development efforts more relevant and effective. Objectivity – truth and impartiality – is valued in participatory evaluation as well as in other types of evaluation, but independence is not seen as one of its necessary preconditions. As it allows participants to engage in open and disciplined reflection on questions concerning the

public good, participatory evaluation can also be seen as a kind of self-education in democratic governance.

It is often useful to distinguish between a general concept of stakeholder participation that includes participation by program staff, government officials, donor representatives, etc. as well as participation by members of target groups, and a narrower concept of so-called popular participation focusing on the participation of target group members. As the term is often understood, a participatory evaluation is not just any evaluation where stakeholders actively participate, but specifically one where primary stakeholders participate directly and in depth.

The best way to promote participation in evaluation – popular participation as well as stakeholder participation more generally – is probably to ensure that a participatory approach is adopted from the very beginning of the intervention process. Introducing a participatory approach towards the end is less likely to be suc-

Box 6

Expert evaluation and participatory evaluation

	EXPERT EVALUATION	PARTICIPATORY EVALUATION
WHAT	Information required by funding agencies and other external stakeholders.	To empower participants to initiate, control and take corrective action.
WHO	External expert evaluators in consultation with stakeholders.	Community members in collaboration with project staff and external facilitators.
WHAT	Standards of performance externally defined, often with reference to formal goals and objectives.	Community members and other participants set their own standards of success.
HOW	External evaluators control data gathering and analysis. Scientific criteria of objectivity. Outsider perspective. Long feedback loops.	Self-evaluation. Collaborative processes of data collection and analysis. Simple qualitative and quantitative methods. Immediate sharing of results.
WHEN	Mid-term, completion, ex-post	Continuous and iterative. Not sharply distinguished from monitoring

Adapted from Deepa Narayan, Participatory Evaluation. World Bank, 1993.

cessful. The following are often mentioned as preconditions for the successful use of participatory approaches to evaluation in community-based projects and programmes:

- Shared understanding among beneficiaries, programme staff and other stakeholders of programme goals, objectives and methods.

- Willingness among programme partners to allocate sufficient time and resources to participatory monitoring and evaluation.

- Participatory approach to programme management and learning. Adopting a participatory approach to evaluation can be difficult if the intervention has been planned and managed in a top-down fashion.

- A reasonably open and egalitarian social structure. Where local populations are internally divided by distinctions of class, power and status a participatory approach is not likely to be successful.

Evaluations initiated by Sida are usually expert evaluations where Sida in consultation with partner organisations formulate the questions and externally recruited expert evaluators provide the answers. Although the evaluators may be instructed to consult closely with target groups and other primary stakeholders during the process, the evaluation agenda is defined by the evaluators – or by the terms of reference for the evaluation – rather than by the primary stakeholders affected by the reviewed activities, as would have been the case in a participatory study.

Still, although Sida rarely initiates participatory evaluations, it may have an interest in promoting such evaluations by the organisations that it helps support. Every development programme is expected to have some kind of system of monitoring and evaluation, and for the donor the design of that system is not unimportant. While not seeking to impose on their partners any particular format for monitoring and evaluation, Sida would normally encourage the adoption of a format consistent with principles of stakeholder participation and responsiveness to citizens.

1.7 Evaluation partnerships

Evaluations of development interventions are carried out in a context of partnership where the government and people of the developing country are the effective owners of the activities encompassed by the relationship and the role of the external donors is a more restricted one of providing financial and technical support. The relationship is based on shared objectives, mutual accountability, and dialogue.

The concept of development partnership has obvious implications for evaluation. Normally, some provisions for evaluation would be included in the partnership agreement itself. The agreement would at least indicate when and why evaluations should be carried out.

According to the established model for development partnerships, evaluations should be managed by the partner country organisation. In practice, however, this task usually falls on the donor. Lack of evaluation capacity in many developing countries is an important reason for this. There is also a tradition on both sides of the development co-operation relationship of regarding evaluation as primarily a tool for donor control, implicitly or explicitly. On the one side, we find donors evaluating for their own purposes, forgetful of the interests of their partners. On the other side we find recipients for whom an evaluation represents a threat or a nuisance rather than an opportunity.

To achieve a real partnership in evaluation, developing countries must strengthen their

capacities for monitoring and evaluation and seek to create political and administrative cultures that are conducive to evaluation and results-based styles of management. Donors, on the other hand, must support those efforts, and at the same time seek to reform existing practices of monitoring and evaluation so that they better serve the interests of their partners. On both sides, learning new ways of organising and conducting evaluations go hand in hand with unlearning practices that have become dysfunctional and obsolete.

In many respects the process of change is already well underway. An important development is the partial shift in the architecture of development co-operation from a fragmented system of project support to a more integrated system of programme support where donors align their contributions with partner country policies and seek to harmonise procedures between themselves. This change affects monitoring and evaluation along with everything else. In sector-based programmes, where the inputs of several donors cannot be separately evaluated, various forms of joint evaluation are becoming the norm (Box 5). For the time being, joint evaluations tend to be dominated by donors, but as the evaluation capacity of partner countries grows the balance will change.

While new types of evaluation are emerging, however, established ones will remain in use, although perhaps in modified forms. Like many development agencies, Sida is still a project-based organisation to a large extent, and this is not likely to change radically in the short run. This means that Sida will continue to make project-level evaluations, in collaboration with its partners or alone. Such evaluations provide inputs to decisions about resource allocation and processes of organisational learning. They also produce information that Sida needs for its own accountability purposes.

"Evaluations should be carried out in a spirit of partnership"

It is Sida policy that evaluations should be carried out in a spirit of partnership. When Sida initiates an evaluation, the evaluation manager should

---

Box 7

Joint evaluation

An evaluation that is sponsored by several donor agencies and/or partner organisations is known as a joint evaluation. It is useful to distinguish between joint multi-donor evaluations and joint donor-recipient evaluations. An evaluation of the latter type can involve one or several donor organisations along with one or more organisations from partner countries. A joint evaluation can be methodologically conventional, and is not the same thing as a so-called participatory evaluation, where primary stakeholders and others are actively involved in the evaluation research process.

Understandably, joint evaluations tend to focus on matters of common interest to their sponsors. A joint evaluation, including both donors and recipients, is a suitable format for assessing sector-wide approaches (SWAps) and other programmes where the contributions of different participating organisations cannot or should not be separated from each other. Interventions that are supported by a single donor organisation are normally best evaluated by that organisation in collaboration with the concerned partner country organisation.

Joint evaluations are used for learning as well as for accountability, although they are perhaps especially useful when the evaluation purpose is to identify good practice and draw useful lessons for the future. Since a joint evaluation builds on the experiences of several organisations, it is likely to have a wider and, in some cases, more powerful impact than an evaluation commissioned by a single organisation.

A useful discussion about joint multi-donor evaluations is provided in the OECD/DAC paper Effective Practices in Conducting a Multi-Donor Evaluation. Paris, 2001.

invite partner country organisations to partici-
pate in the evaluation process. In some cases, the
partner organisation will prefer not to become
actively involved; in other cases it will welcome
the invitation. Questions that tend to be impor-
tant to donors are often just as important to their
co-operation partners, and vice versa. The results
of jointly supported interventions are clearly a
matter of common concern. Questions about
aid effectiveness – the effects of aid as opposed
to the effects of the interventions that aid helps
finance – and the development co-operation
relationship itself should also be important to
both the parties.

Note that donors should avoid embarking on
evaluations without first assessing the utility of
those evaluations in relation to their cost to part-
ner country organisations. Where there are many
donors, each with its own set of projects or pro-
grammes, evaluations commissioned by donors
can be a burden on the weak administrative
systems of developing countries. Among donors
pre-occupied with the success of 'their' projects,
this is easily overlooked. One of the arguments
in favour of closer co-operation in evaluation –
between donors as well as between donors and
their developing country partners – is that it
would reduce the transaction costs of aid and give
overburdened partner country organisations more
time to deal with other important matters.

1.8 Quality standards

Evaluation is a tool for quality assurance and
quality control, and as such must satisfy its
own quality requirements. In this manual
we suggest that the quality of any particular
evaluation can be assessed in relation to four
broad sets of quality standards, namely those of
propriety, feasibility, accuracy and utility. [7]

The *propriety standards* are ethical standards
meant to ensure that evaluations are conducted
with due regard for the rights and welfare of
affected people. The most basic of the propriety
standards is that evaluations should never violate
or endanger human rights. Evaluators should
respect human dignity and worth in their inter-
action with all persons encountered during the
evaluation, and do all in their power to ensure
that they are not wronged.

The principle of *informed consent,* which is cen-
trally important to the ethics of scientific research,
is also relevant to evaluation. The twofold require-
ment of this principle is that a) people should not
be engaged as respondents, informants or parti-
cipants in evaluations without their consent, and
b) people should be given adequate information
about the evaluation, its purposes and possible
consequences, before they are actively involved.

It is an important aspect of propriety that
evaluations should be balanced and fair, both in
the research phase and in the final stage of
reporting. All relevant stakeholder groups should
be allowed to speak out and their views should
be correctly reported. People have a right not to
be misrepresented. Furthermore, evaluations
should normally be concerned with the strengths
and weaknesses of systems, structures, and forms
of organisation, rather than with the strengths
and weaknesses of particular individuals or
groups of individuals.

The *feasibility standards* are intended to ensure
that evaluations are realistic and efficient. To
satisfy these requirements, an evaluation must be

---

[7] These are the main standards of the Program Evaluation Standards of the American Joint Committee on Standards for
Educational Evaluation (1994). A summary of these standards can be downloaded from www.wmich.edu/evalctr/jc/.

based on practical procedures, not unduly disrupting normal activities, and be planned and conducted in such a way that the co-operation of key stakeholders can be obtained. They should also be efficient. If the cost of an evaluation cannot be justified by the usefulness of the results to intended users, it should not be undertaken.

The *accuracy standards* are meant to ensure that the information produced by evaluations is factually correct, free of distorting bias, and appropriate to the evaluation issues at hand. By setting high standards for accuracy, we protect the very function of evaluation as a means of making sure that plans and expectations are based on reality and not the result of prejudice or wishful thinking.

> "The quality of any particular evaluation can be assessed in relation to four broad quality standards; those of propriety, feasibility, accuracy and utility"

However, it must be recognised that accuracy is not an end in itself. As evaluations can be costly and time-consuming for all parties involved, efforts to achieve a high level of accuracy should be tempered by a pragmatic principle of "optimal ignorance." In others words, we should not strive for the highest degree of accuracy possible, but for one that is good enough for the purpose at hand and can be accepted as such by the users of the evaluation.

The *utility standards*, finally, are meant to ensure that evaluations serve the information needs of their intended users. An evaluation that users consider irrelevant is hardly a success, regardless of its other merits. To be useful, evaluations must be responsive to the interests, perspectives and values of stakeholders. It is important that evaluations are *timely* in relation to stakeholders' practical agendas, and that stakeholders regard them as *credible*.

The credibility of an evaluation depends on several factors. First, the evaluators must be accepted as impartial and unbiased. If the evaluation appears to be influenced by partisan interests, stakeholders will reject it. Second, the evaluators must be technically and culturally competent to deal with the questions raised by the evaluation. If they are thought to lack necessary qualifications, the credibility of the evaluation is again compromised. Third and finally, methods and resources for data collection and analysis must be regarded as appropriate. An evaluation that passes the other tests, but fails in this one, may still be rejected.

Chapter 2

An evaluation is an assessment of the merits and worth of a project, programme or policy in relation to a particular set of evaluation criteria and standards of performance. Defining criteria and standards is a key step in every evaluation. This chapter explains the distinction between criteria and standards and presents five criteria that are particularly important in assessments of development of interventions:

- Effectiveness
- Impact
- Relevance
- Sustainability
- Efficiency

Used by development organisations around the world, these criteria are essential components of development co-operation evaluation. The chapter has a separate section for each one of them. Each section concludes with a set of standard evaluation questions.

Evaluation Criteria

2.1 Principal evaluation criteria

Every evaluation involves one or several criteria by which the merit or worth of the evaluated intervention is assessed, explicitly or implicitly. The following five have been recommended by the OECD/DAC and adopted by Sida as standard yardsticks for the evaluation of development interventions:

EFFECTIVENESS
The extent to which a development intervention has achieved its objectives, taking their relative importance into account.

IMPACT
The totality of the effects of a development intervention, positive and negative, intended and unintended.

RELEVANCE
The extent to which a development intervention conforms to the needs and priorities of target groups and the policies of recipient countries and donors.

SUSTAINABILITY
The continuation or longevity of benefits from a development intervention after the cessation of development assistance.

EFFICIENCY
The extent to which the costs of a development intervention can be justified by its results, taking alternatives into account.

Each one of these criteria can be applied to every development intervention and each one of them represents something important that needs to be considered before it can be decided if a particular intervention should be regarded as a success.

The best way to get a good understanding of the five criteria is to start with the distinction between effectiveness and efficiency. What we should bear in mind is simply that these two are fundamentally different. Effectiveness refers to the extent to which an evaluated intervention has achieved its objectives, nothing else. Efficiency, by contrast, refers to the extent to which the costs of an intervention can be justified by its results. An analysis of efficiency makes little sense without a prior assessment of the results.

> **A point about language**
> Swedish-speakers often find the distinction between effectiveness and efficiency awkward and confusing. The reason for this is that the Swedish word *effektivitet* covers both efficiency and effectiveness. In many situations it refers to efficiency alone. In technical language, the word *kostnads-effektivitet* – cost-effectiveness – serves as an equivalent of efficiency. Still, for Swedes consistently separating efficiency from effectiveness may require a special effort.

The remaining criteria – impact, relevance, and sustainability – complement the criterion of effectiveness. Like the latter, they are used to assess results independently of costs. With impact, we transcend the managerial bias inherent in assessments of effectiveness and attempt instead to examine the evaluated intervention from the point of view of target groups and other primary stakeholders. The concept of unintended consequences is centrally important in impact studies. When we focus on relevance, we look at the evaluated intervention in relation to larger contexts of needs, priorities, and policies, and we examine the objectives of the intervention as well as the means of achieving them. Sustainability, finally, is a criterion for assessing the likelihood that the benefits produced by an intervention will be maintained beyond the cessation of external support.

We should not be confused by the fact that the five criteria tend to overlap at several points. For example, in a study of impact we will encounter some of the same effects as we may already have dealt with in an assessment of effectiveness. Similarly, there may be overlaps between a study of relevance and a study of effectiveness. However, this becomes a problem only if we expect the five criteria to be concerned with completely different sets of facts, which is not the case. Each criterion represents a particular way of looking at an intervention, nothing else.

The five criteria are discussed at length in sections 2.4–2.8.

2.2 Further criteria

As stated in Sida's Evaluation Policy, all the five standard criteria should be considered when Sida initiates an evaluation. An evaluation using all of them would give us much of the information needed in order to form an overall opinion of an intervention's value. Still, the policy does not require all five to be adopted in each and every case. The policy requirement is rather that none of them should be put aside without a prior assessment of their relevance.

Furthermore, although these criteria have been accorded a special status by Sida's Evaluation Policy, we are not prevented from using additional criteria. In many evaluations procedural values and principles are used as evaluation criteria. Participation, partnership, human rights, gender equality, and environmental sustainability, are prominent examples. They are all values and principles governing the design and implementation of interventions supported by Sida, as well as major policy goals. It would be surprising if they were not also used as criteria for evaluation. Participation is discussed in 3.6, the gender equality criterion in 3.7, and the environmental perspective in 3.8.

Note also the existence of evaluation criteria related to particular areas of development co-operation. In research co-operation, for example, the term relevance can refer both to the scientific usefulness of a research project and to the relevance of such a project in a general development perspective. Other examples are the criteria of appropriateness, coverage, connectedness, and coherence that have been designed to fit the special conditions of humanitarian assistance. Sponsored by the OECD/DAC these criteria are regarded as subcriteria to the five principal criteria discussed in this chapter. Definitions and brief explanatory notes are given in Box 8.

Criteria for evaluation of humanitarian assistance

APPROPRIATENESS

The extent to which humanitarian inputs and activities are tailored to local needs, and the requirements of ownership, accountability, and cost-effectiveness. How well did the humanitarian activities respond to the changing demands of the situation?

COVERAGE

The extent to which the entire group in need had access to benefits and were given necessary support. Key questions: Did the benefits reach the target group as intended, or did too large a portion of the benefits leak to outsiders? Were benefits distributed fairly between gender and age groups and across social and cultural barriers?

CONNECTEDNESS

The extent to which short-term emergency activities take into account longer-term needs and the interconnected-

ness of humanitarian problems. Examples of problems to be dealt with are environmental effects of refugee camps, damage of roads through food transports, damage to local institutions as a result of international NGOs taking over central government functions and recruiting skilled staff from local government institutions.

COHERENCE

Consistency between development, security, trade, military and humanitarian policies, and the extent to which human rights were taken into account. Important questions: Were policies mutually consistent? Did all actors pull in the same direction? Were human rights consistently respected?

These criteria are extensively discussed in Evaluating Humanitarian Action. An Alnap Guidance Booklet. www.alnap.org

2.3 Performance standards

It is useful to make a distinction between evaluation criteria and performance standards related to those criteria. While evaluation criteria are variables in terms of which performance is measured or assessed, performance standards are values on those variables representing acceptable levels of achievement. For example, if we are dealing with a road construction intervention and effectiveness is the criterion of evaluation, one kilometre of asphalt road completed every three months could be a performance standard.

While performance standards are often taken directly from the definition of objectives and performance targets in funding proposals and other documents relating to the evaluated intervention, there are also other possibilities. Interventions may be assessed against policy-level goals and principles, or they may be evaluated against standards derived from comparisons with similar interventions elsewhere or from models of best practice. An evaluation of an intervention in relation to Sida's gender policy would be an example of the first type of assessment. An evaluation of the same intervention in relation to what is internationally regarded as best practice for promoting gender equality would be an example of the second.

In some cases, standards of performance must be defined during the evaluation process itself, rather than during planning. Furthermore, even if the standards of performance are well defined in

advance the evaluators may have to decide whether they are reasonable or not, given the problem addressed by the intervention and the amount of resources invested. When looking at the facts of the case, the evaluators may come to the conclusion that the performance standards were too ambitious or not ambitious enough. They might even propose an alternative set of standards. Other complications regarding standards of performance are due to the fact that the goals and objectives of development interventions may be vague, incomplete or both. Problems of this kind are briefly discussed in 2.4.

Regardless of how standards of performance are defined, however, it is essential that they reflect the goals, commitments and formal obligations of those who are involved in the evaluated activity and that they are not regarded as arbitrary impositions. Standards of performance that evaluators or other outsiders define ex post according to their own predilections, insiders may regard as unfair or irrelevant.

2.4 Effectiveness

The term effectiveness refers to the extent to which the objectives of an intervention have been achieved as a result of the implementation of planned activities. Effectiveness can be measured at the level of outputs as well as at the levels of outcome and impact. In the first case we are concerned with the achievement of targets for the production of goods and services, in the second with the achievement of the further effects that we intend to bring about through these goods and services.

Note that effectiveness at the output level is no guarantee for effectiveness in terms of outcomes and impacts. An intervention may achieve all its targets with regard to goods and services, and still not be effective at the outcome and impact levels. Indeed, an intervention that is effective in terms of outputs may be quite ineffective with regard to outcomes and impact. An important implication of this is that every question about effectiveness in evaluations should be carefully specified in relation to the intended level of objectives. Effectiveness in general is an empty concept.

Assessing effectiveness is usually more difficult at the level of outcomes and impact than at the output level. At the output level, the job can be regarded as completed when we have measured the extent to which the goods and services produced by the intervention match pre-defined targets for quantity and quality. Assessing the quality of outputs can be difficult, especially where clear quality standards are lacking. In most cases, however, there are solutions to such problems.

At the outcome level an assessment of effectiveness is made in two steps. First, the *achievement of objectives* is measured. How have the conditions of the target group changed since the intervention was launched and how do identified changes compare with intended changes? Second, the issue of *causal attribution* is addressed. To what extent have the identified changes been caused by the intervention rather than by factors outside the intervention?

Both these steps are important. For a government or development agency that wishes to invest its resources where they make a difference, knowing that things have changed as expected is not good enough. They also want to know that the evaluated intervention has significantly contributed to the recorded change. In most cases there are many factors at play in addition to the evaluated intervention. Therefore, the possibility

that recorded changes would have occurred even without the intervention must be carefully considered, and as far as possible ruled out.

Measuring goal achievement is, in principle, a fairly uncomplicated affair. In practice, however, it may be very difficult or even impossible. The reason for this is that one or several of the following types of information may be missing:

- Baseline information about the situation before the intervention (see Box 9).
- Records of the changes that have occurred relative to the baseline during the implementation of the intervention and later.
- An empirically verifiable description of the state of affairs that the intervention was intended to bring about.

Missing baseline information is a common problem. In some cases a baseline can be reconstructed with the help of written documents and interviews with members of target groups and others. Human memory being what it is, however, a baseline reconstructed through memory recall is usually much less precise and much less reliable than a baseline assembled before the intervention started.

Vagueness in the description of the state of affairs that the intervention is intended to bring about can also cause problems. If goals and objectives are unclear it may be difficult to decide if and to what extent recorded changes represent an achievement in relation to them. In many interventions goals and objectives are further specified through empirical indicators, but indicators may raise questions of their own. As evaluators, we should never take for granted that the indicators are valid and relevant.

Formulations of objectives tend to become increasingly vague as we move upwards in the hierarchy of objectives. At the level of outputs it is often quite clear what the intervention intends to achieve and at the level of immediate outcomes the objectives may also be reasonably precise. In the higher reaches of the goal hierarchy, however, the descriptions of objectives are often extremely vague. Often no stronger or more precise claim is made than that the intervention is expected "to contribute" to the achievement of a distant higher goal, or that it will help "create pre-conditions" for the higher goal to be achieved. Measuring goal achievement in relation to objectives that are formulated in such terms is difficult, if not altogether impossible.

"The possibility that recorded changes would have occurred even without the intervention must be carefully considered"

Despite these difficulties, however, measuring goal achievement is often a lesser problem than attributing causality. There are cases where evaluators do not hesitate to attribute the recorded change to the intervention. In other cases, however, causal attributions are much less certain. Usually, short-term changes are more strongly influenced by the evaluated intervention than longer-term changes. Conversely, the more distant the change the more likely extraneous factors are to intervene. In many cases, we can make no stronger claim regarding causality than that the intervention appears to have had some influence on the recorded outcomes.

Section 2.5.2 on page 32 contains some further remarks on the problem of inferring causality.

Effectiveness is an important criterion for the evaluation of development interventions. It is important to principals, financiers and others who for reasons of accountability want to know that results have been delivered as promised. It is also important to intervention managers who need the same type of information for management and learning. Setting goals and defining objectives that are both challenging and realistic is a basic management skill. If we never reach our goals and never achieve our objectives something is clearly wrong, even if we can rightly claim to have produced results that are valuable and worthwhile.

Box 9

What is a baseline study?

A baseline study is a description of conditions in a locality or site prior to a development intervention. A baseline study provides benchmarks against which change and progress can be measured and evaluated. Without baseline information, assessments of effectiveness and impact are impossible. Baseline information can often be assembled retrospectively, but, as a rule, a reconstructed baseline is much inferior to baseline information assembled ex ante.

The scope and focus of a baseline study reflect the purpose of the intervention and the anticipated future need for baseline data in monitoring and evaluation. If the intervention is a project in support of entrepreneur-

ship among small farmers, for example, the baseline study will describe the nature and extent of rural entrepreneurship before the intervention. If it has a subsidiary goal of increasing the number of female entrepreneurs, the baseline data should be disaggregated by gender. If it is expected that the project may affect the physical environment, the baseline will contain benchmarks for the monitoring and evaluation of environmental effects.

Useful reading: Soleveigh Freudenthal and Judith Narrowe. Baseline Study Handbook. Focus on the Field. Sida, Stockholm, 1993.

By itself, however, an assessment of effectiveness has only limited value. What we can learn from it is whether the evaluated intervention has achieved its goals and objectives, nothing else. As already noted, an assessment of this kind says nothing about the relevance and value of the achieved results; neither does it provide any information concerning the important question of unintended effects.

Standard questions about effectiveness:

- To what extent do development changes in the target area accord with the planned outputs, purpose and goal of the evaluated intervention?
- To what extent is the identified development the result of the intervention rather than extraneous factors?
- What are the reasons for the achievement or non-achievement of objectives?
- What can be done to make the intervention more effective?

2.5 Impact

The word impact has several meanings. It often refers to the totality of the effects brought about by an intervention, but it is also used more narrowly to refer to effects in the longer term or to effects at the scale of societies, communities, or systems. When it is used in the narrow sense – as in logframe analysis for example – it is complemented by the word outcome, which refers to short and medium term effects on the attitudes, skills, knowledge, or behaviour of groups or individuals. In other contexts, the term outcome may refer to the totality of effects produced by an intervention, just as impact.

In this manual the term impact is used in both the senses. This should not create any problems. When used in conjunction with the term outcome, it refers to effects in the longer term. Elsewhere it means the totality of effects brought about through a development intervention. In the latter sense it encompasses expected and unexpected, positive and negative, as well as short-term and long-term effects on people, organisations, societies and the physical environment.

A study of impact in the wide sense covers partly the same ground as a study of effectiveness. However, it differs from such a study in two important respects. First, while assessments of effectiveness may deal with outputs as well as effects, an impact study is limited to effects. Second, while studies of effectiveness tend to focus on planned positive effects in the short or medium term, a study of impact is concerned with the entire range of effects, including those that were unforeseen, those that occur in the longer term, and those that affect people outside the target group.

Note that impact is not a criterion of the same kind as effectiveness or efficiency. While the latter are normative criteria – effectiveness and efficiency are by definition desirable – the impact criterion is primarily a recommendation that all significant consequences of an intervention, negative as well as positive, should be taken into account. To be able to distinguish between positive and negative impacts we have to employ additional normative criteria, such as increased well being of primary stakeholders, utility for poverty reduction, or something of that kind.

The impact criterion provides an important corrective to what could otherwise become an overly narrow preoccupation with the intentions of those who plan and manage development interventions and a corresponding neglect of the perspectives of target groups and other primary stakeholders. This is a key point about impact. When applying the impact criterion, we turn to target groups and other stakeholders to find out if and how the evaluated activities have affected their situation, positively or negatively. Measuring change in relation to stated intervention objectives is less important.

Evaluations that give precedence to the perspectives of target groups and other primary stakeholders over those of planners and managers are sometimes referred to as *user-oriented evaluations* (in contrast to *goal-oriented evaluations* focusing on effectiveness). The term *goal-free evaluation* has a similar meaning. In a goal-free evaluation the evaluators deliberately try to disregard the established goals and purposes of an intervention in order that they may better appreciate the value and significance of the intervention to those who are affected by it.

2.5.1 Types of impact

As shown in Box 10, a study of impact deals with four types of effect. In the left-hand column there are the intended positive effects justifying the intervention and the negative effects that those who are responsible for the intervention anticipate and accept as affordable costs or necessary evils. The right-hand column contains the unexpected effects, positive as well as negative.

Box 10

Types of impact

EXPECTED POSITIVE	UNEXPECTED POSITIVE
EXPECTED NEGATIVE	UNEXPECTED NEGATIVE

Evaluations should cover both positive and negative effects. In the planning of development interventions possible negative effects are easily overlooked or taken too lightly. Truly unexpected negative effects are caused by ignorance, bad planning or wishful thinking. Normally, stakeholders and outside expertise are extensively consulted when development interventions are planned but even so things may not turn out quite as expected. In some cases, the actual results may even be the opposite of the intended results. Identifying unintended effects and analysing their causes is one of the main tasks of evaluation.

Evaluators have the advantage of hindsight. In many cases, they can identify unintended consequences simply by asking primary stakeholders to describe how their situation has changed since the intervention started. Although identifying impact in this way can be time-consuming and may require a good deal of local knowledge, it need not be technically complicated. Still, interviewing representatives of target groups and other stakeholders is not a universal recipe for impact study. Interventions may have adverse effects that only experts can identify. Negative effects that are difficult to identify for laypersons can be found in all areas of development co-operation. Box 11 lists a few common types of negative unintended effects.

2.5.2 Measuring change and inferring causality

When studying impact we face the same technical problems of measuring change and inferring causality as in studies of effectiveness.

However, the first of the two tasks, that of measuring change, can be rather more difficult in impact studies than in assessments of effectiveness. As noted in 2.4, change cannot be measured without a baseline describing the situation before the intervention (Box 9, page 30). Since baseline information assembled before the intervention is likely to refer to those features of the situation that are expected to change as a result of the intervention, however, it is not always useful for measuring change that was not planned and not foreseen. To be able to measure unexpected change, we may have to construct a new baseline ex post. However, in many cases constructing a baseline after the fact may be difficult or even impossible.

The second main task is to decide, with as much certainty as required or possible, whether the changes that have occurred since the beginning of the intervention were caused by the intervention, or if they would have occurred anyway. Impact, in the strict sense, is the difference between the changes that have actually occurred and the changes that *would* have occurred without the intervention. The hypothetical state of affairs to which we compare real changes is known as the *counterfactual*.

With the help of *control groups* that have not been exposed to the intervention it is sometimes possible to get a good idea of how the target group would have fared without the intervention. When the counterfactual cannot be estimated in this way – a common situation in development co-operation – statements about impact rest on weaker foundations. The intervention is often taken to be the cause of the identified changes if such a conclusion appears to be consistent with expert knowledge and there seems to be no better explanation around. Although less compelling than an explanation based on control group methodology in most cases, an argument of this type can be good enough for the purpose of the evaluation.

For a deeper discussion about causal attribution in relation and different types of research design the reader should consult a textbook on evaluation research methods.[8]

[8] See, for example:
  Bamberger, M. and Valadez, I. (eds.). Monitoring and Evaluating Social Programs in Developing Countries. Washington D.C: The World Bank, 1994.
  Rossi, P. et al. Evaluation. A Systematic Approach. Seventh Edition. Thousand Oaks: Sage, 2004.

Box 11

Types of negative unintended effects

TARGETING ERRORS

There are two main types, often referred to as Type I and II Errors. The first type is concerned with coverage: Have benefits from the intervention reached all sub-groups of the targeted category, or is there a coverage bias such that some of the intended beneficiaries – women, children, the elderly, or the disabled, for example – have been excluded? The second type is concerned with leakage of benefits to people outside the targeted category. In other words, the question is whether outsiders have seized benefits intended for the target group. Is there a problem of overinclusion?

SUBSTITUTION AND DISPLACEMENT

In both cases the intended positive effects for a particular target group are realised, but only at the expense of another group or category that is equally deserving of support. A case of substitution would be that of sub-sidised workers replacing unsubsidised workers who would otherwise have been employed. Displacement, on the other hand, would occur if the creation of sub-sidised jobs in one firm led to reduction of workers in other firms.

RECOIL EFFECTS

These are unintended effects of an intervention on the organisations responsible for its implementation and management. The aim of development assistance is not just to help solve immediate development problems of a pressing kind, but to strengthen the capacity of developing countries and their organisations to deal with such problems with their own resources. In many cases, however, recipient country organisations are overburdened by externally financed development interventions. Evaluations can be used to assess the extent to which development initiatives produce such negative recoil effects.

FUNGIBILITY

By lending support to a particular activity the donor makes it possible for the recipient to shift scarce resources to other activities, all or some of which may be inconsistent with the donor's mandate. Thus, while the donor focuses on the activities ear-marked for sup-port, the main impact of aid is perhaps to make possi-ble activities that the donor regards as undesirable. Important as it may be, however, the phenomenon of fungibility can usually not be handled within an evalua-tion of a single development intervention. To find out if and how donor support is fungible, it must be examined in relation to the national budget and wider patterns of public expenditure in the partner country. Note that it is only when the recipient's priorities are inconsistent with those of the donor that fungibility becomes a problem.

PERVERSE EFFECTS

Substitution and displacement, fungibility, as well as some of the recoil effects of aid programmes can all be described as perverse effects. Effects referred to by this term are directly opposed to the objectives and goals that the intervention was intended to achieve. Donor support that is meant to strengthen the ability of people to deal with their own problems but results in increased dependency on aid is a prime example from development co-operation.

Standard questions about impact:

- What are the intended and unintended, positive and negative, effects of the intervention on people, institutions and the physical environment? How has the intervention affected the well being of different groups of stakeholders?

- What do beneficiaries and other stakeholders affected by the intervention perceive to be the effects of the intervention on themselves?

- What is the impact of the intervention on the recipient country organisation(s) that manage it? To what extent does the intervention contribute to capacity development and the strengthening of institutions?

- To what extent can changes that have occurred during the life span of the intervention or the period covered by the evaluation be identified and measured?

- To what extent can identified changes be attributed to the intervention? What would have occurred without the intervention?

- Have plausible alternative explanations for identified changes been considered and convincingly ruled out?

2.6 Relevance

When we talk about relevance as a criterion of evaluation, we are concerned with the value and usefulness of the evaluated intervention in the perspectives of key stakeholders. More precisely, a development co-operation intervention is considered relevant if it matches the needs and priorities of its target group, as well as the policies of partner country governments and donor organisations.

The question of *needs* comes first. If an intervention does not help satisfy important development needs, directly or indirectly, it can obviously not be regarded as relevant. However, the fact that it addresses important needs is not enough. To be considered relevant a development intervention should also be consistent with the *policies* and effective *priorities* of target groups and others. In addition, it should be *technically adequate* to the problem at hand – an effective and efficient cure without unwanted side effects, as it were. The latter requirement is implied by the definition. If the intervention is out of tune with stakeholder priorities, or if it is technically inadequate in some way, it will probably not achieve its aims, and, again, is not relevant as a solution to the development problem at hand.

In many evaluations, the objectives of the evaluated intervention are taken as given. When the relevance of an intervention is assessed, however, the objectives as well as the means of achieving them are carefully examined. The perspective is holistic. At one level, we try to ascertain if the intervention is well adapted to the livelihood patterns and the social and political conditions of its intended end-users and other primary stakeholders. At another level, we wish to establish that it is well in line with government policies and systems of administration as well as with concurrent interventions supported by other development agencies.

Questions about partner country ownership are important in assessments of relevance. To what extent was the evaluated intervention an independent host country initiative; to what extent was it rather an opportunistic adaptation to donor preferences? To what extent is it managed by host country actors? To what extent are those actors prepared to invest their own resources in it? Does ownership extend to the intended beneficiaries and other citizens? Are there adequate

mechanisms for accountability and popular participation?

While stressing that relevance is about consistency with existing priorities and policies – effective demand – as well as needs, however, we should not lose sight of the experimental nature of many development interventions. The point is obviously not that interventions that challenge established interests or existing ways of doing things are always irrelevant. It is rather that even interventions that go against the grain of existing practice must be well grounded in effective interests and priorities. When we assess the relevance of an innovation, one of the key questions is often whether it has a potential for replication.

Relevance is an important issue throughout the intervention cycle. At the stage of planning and preparation, the responsible organisations make a first assessment of the relevance of the objectives of the intervention, and they also try to make sure that the intervention strategy is sound. Later, interim or ex-post evaluations should revisit this analysis. The initial assessment may have been incorrect all along, or the situation may have changed in such a way that it has to be revised.

Standard questions about relevance:

- Is the intervention consistent with the livelihood strategies and living conditions of its target group? How urgent is it from the point of view of the target group?

- Is the intervention well in tune with the development policies and administrative systems of the partner country government at national and regional levels? Is it consistent with a policy of supporting partner country ownership?

- Is the intervention a technically adequate solution to the development problem at hand? Does it eliminate the main causes of the problem?

- Do proposed innovations have a potential for replication?

- Is the intervention consistent with Sida policies and priorities?

- Is the intervention consistent and complementary with activities supported by other donor organisations?

2.7 Sustainability

When we discuss sustainability we are concerned with the likelihood that the benefits from an intervention will be maintained at an appropriate level for a reasonably long period of time after the withdrawal of donor support. Strictly speaking, sustainability is part of the wider criterion of impact, but as it concerns a recurrent issue of extreme importance in development co-operation we treat it as an independent criterion.

The criterion of sustainability refers to the results obtained through development co-operation interventions, not the development co-operation interventions themselves. In some cases, sustainability means that a particular organisation or facility constructed with external assistance will remain in use, but in other cases the organisation or facility built in the course of the intervention more resembles a temporary scaffolding that is needed only in the construction phase.

Sustainability must be specified in relation to the particular intervention under review. Different types of intervention have different time frames and serve different types of function in the development process. Sustainability does not have the same meaning in regard to short-term emergency assistance as in interventions with

long-term development objectives. While the evaluation criterion is the same in both cases, the performance standards will differ.

Analyses of sustainability in evaluations are forward-looking assessments made during the implementation process or when the intervention has been completed. The main question that such analyses seek to answer is usually not whether intervention benefits have in fact been sustained. This is a question that in most cases cannot be answered ahead of time. The question is rather if the intervention has a potential for being sustained, and if it is likely that its positive impact will be a lasting one.

There is a range of factors that determine whether or not the results of the evaluated intervention will be sustained into the future:

PARTNER COUNTRY PRIORITIES

Development interventions always operate in the policy environment of partner countries. The priorities of partner organisations are critical to the sustainability of their results. Interventions stand a much better chance of being sustained if they reflect partner country priorities than if donors drive them.

PARTNER COUNTRY OWNERSHIP AND PARTICIPATION

Without partner country ownership development interventions can usually not be sustained. The active participation of partner country stakeholders in the planning, implementation and follow-up of development activities helps stimulate local ownership.

INSTITUTIONAL AND CULTURAL FACTORS

Interventions should be well integrated in the local institutional and cultural context. Interventions that are out of tune with local norms and sensibilities, or lack institutional support, are unlikely to be sustained.

TECHNOLOGICAL FACTORS

The technology utilised in the intervention should be appropriate to the economic, educational and cultural conditions of the host country. If the level of technology is too advanced and spare parts scarce or too expensive, continued maintenance is unlikely.

ENVIRONMENTAL FACTORS

In developing countries the natural environment is often under pressure from population growth and poor management of natural resources. Environmental degradation may force partner organisations to discontinue otherwise positive intervention results.

FINANCIAL FACTORS

In many interventions, sustainability depends on partners' financial capacity to maintain results. Government revenue, user fees and other income generating activities may secure such funding, and hence contribute to sustainability.

MANAGEMENT AND ORGANISATION

The governance of recipient country institutions is a key determinant of sustainability. Weak, ineffective, or "unsound" management and organisation may significantly reduce the likelihood that results will be sustainable. In many cases, a working system of accountability to citizens is likely to increase the chances that benefits will be sustained.

EXIT STRATEGY

The chances that the benefits from an intervention will be sustained are also likely to increase if there are time limits and well defined exit points. An exit strategy, including a plan for sustainability, should be part of every development intervention from the very beginning.

The above list of factors is not necessarily exhaustive. Neither is it a catalogue, which must be included in all evaluations initiated by Sida. The analysis of sustainability must always be adapted to the circumstances of each evaluation.

Standard questions about sustainability:

■ Is the intervention consistent with partners' priorities and effective demand? Is it supported by local institutions and well integrated with local social and cultural conditions?

■ Are requirements of local ownership satisfied? Did partner country stakeholders participate in the planning and implementation of the intervention?

■ Are relevant host-country institutions characterised by good governance, including effective management and organisation?

■ Is the technology utilised in the intervention appropriate to the economic, educational and cultural conditions in the partner country?

- Do partners have the financial capacity to maintain the benefits from the intervention when donor support has been withdrawn?
- Is the intervention compatible with a sustainable use of natural resources? Or is it harmful to the natural environment?

2.8 Efficiency

Efficiency is a relation between means and ends. More exactly, it is the ratio of the value of the results of an intervention to the value of the resources used to produce them. An intervention is optimally efficient if its value is greater than the value of any alternative use of these resources. If the same resources could have produced better results in some other way, or if the same results could have been produced with fewer resources, it is less than fully efficient.

Economists distinguish between different types of efficiency. At it simplest, efficiency is synonymous with *technical efficiency*. When we say that an intervention is technically efficient we mean that it produces as much as possible of a specified result, given the available resources. Alternatively, an intervention is technically efficient when it produces a planned result with a minimum of resources. In either case, technical efficiency is a matter of avoiding waste and reducing costs. An intervention is technically efficient when its productive capacity is fully utilised.

A more complex concept of efficiency is that of *allocative efficiency*. To be efficient in this sense an intervention must first of all be technically efficient. But this is not enough. Optimal allocative efficiency is not achieved unless the intervention produces the highest possible level of utility or value to society. If an alternative use of the same resources would have produced a higher level of aggregate value or utility, the intervention is not optimally efficient from an allocative point of view, although it may well be highly efficient in the technical sense.

It is important to understand the difference between the two concepts. In an assessment of technical efficiency, an intervention is evaluated against other ways of achieving the same concrete objective, regardless of the value of that objective. In an assessment of allocative efficiency, by contrast, an intervention is evaluated against every alternative use of the same resources, at least in principle. Whereas an assessment of technical efficiency focuses on the relation between inputs and outcomes (or outputs) and takes the objective as given, an assessment of allocative efficiency raises the more far-reaching question of whether the intervention is economically worthwhile, given the alternatives foregone.

Assessments of efficiency are known as *economic evaluations*. By a standard definition, an economic evaluation is a comparative analysis of alternative courses of action in terms of both their costs and their consequences. The object of comparison is usually another intervention, but it can also be a constructed model of best practice or standard of performance.

The following are the most important types of economic evaluation: cost-effectiveness analysis (CEA), cost-utility analysis (CUA) and cost-benefit analysis (CBA). All three are forms of assessment where an analysis of outputs and effects like those discussed in previous sections is coupled to an analysis of costs. The most important of the differences between them concern the analysis of the outputs and effects. In CEA results are measured in terms of natural units: number of households supplied with fresh water, number of persons cured of a particular disease, etc. In CBA results are measured in a variety of ways but

valued and compared in monetary terms. In CUA, finally, results are reduced to a generic, non-monetary measure of utility that is applicable to a particular type of intervention or policy domain. In the health sector qualitatively different interventions are made comparable through such generic measures of health status as Quality Adjusted Life Years (QALYs) and Disability Adjusted Life Years (DALYs).

Which of these types of evaluation is the most appropriate depends on the nature of the evaluation questions. If we are interested in the technical efficiency of an intervention in relation to alternative ways of achieving the same substantive results, CEA could be the appropriate tool. If the question concerns allocative efficiency, on the other hand, CBA would be a better choice. When we are dealing with allocative efficiency on a societal scale there is indeed no alternative to CBA. CUA is also a method for assessing allocative efficiency, but more limited in scope. It is used for answering questions about resource allocation at the level of a particular sector of the economy when measuring results in monetary terms seems inappropriate. For example, with QALYs health economists can quantify the utility of different types of health care interventions, while avoiding some of the moral problems of measuring the value of human life in monetary terms.

CBA is the most powerful but also the most debated of the different methods of economic evaluation. In a CBA the evaluator first seeks to identify all the significant costs and benefits of an intervention, direct and indirect, tangible and intangible, and then attempts to estimate the value of each one of them in monetary terms. The costs and benefits may include factors like a healthy or aesthetically attractive environment and human life itself. In the final steps of the analysis, costs and benefits are added up and compared. If the total value of the benefits exceeds the total value of the costs, the intervention is considered worth-while, otherwise not. Costs and benefits that are difficult to measure in monetary terms can be set aside for separate consideration after the completion of the CBA.

CBA is used in prospective appraisals more often than in evaluations of completed interventions. It is intended to facilitate an economically rational use of scarce resources. As both critics and supporters would agree however, CBA is not without limitations. It is often noted that it tends to favour the interests of the present generation over the interests of future generations. Since the value of long-term benefits and costs are calculated from a standpoint in the present – a technique known as discounting– it can lead to an altogether too light-hearted treatment of issues of environmental sustainability and intergenerational equity. When the time-scale of the analysis is restricted to a period of 30–40 years or less, however, there are no such problems.

Another limitation concerns the analysis of distributive effects in the present. In CBA an intervention is regarded as useful and worthwhile if it increases the total amount of satisfaction in society as measured directly by the market or indirectly through consumers' stated 'willingness to pay' elicited by interviews. However, simply aggregating costs and benefits can easily conceal the fact that there may be losers as well as winners. Distributional questions can be added to a CBA, but are not required by the model itself.

What this boils down to is simply that CBA is concerned with economic efficiency rather than equity or social justice. In development co-operation, where we are concerned with poverty reduction and human rights, this is something to bear in mind. Normally, assessments of efficiency provided by CBA, as indeed any efficiency assessment, should be complemented by considerations based on other criteria. The point to remember is that all costs and benefits are relative to particular individuals or groups.[9]

[9] Useful reading on economic evaluation: Belli, P. et al. Handbook on Economic Analysis of Investment Operations. Washington, D.C: World Bank, 1998. Drummond, P. H. et al. Methods for economic evaluation of health care programmes. Oxford: Oxford University Press, 1997. Levin, H. M. Cost-Effectiveness. A Primer. Newsbury Park: Sage, 1983. Rossi, P. H. et al. Evaluation. A Systematic Approach. Seventh Edition. Newsbury: Sage, 2003.

Standard questions about efficiency:

- Has the evaluated intervention been managed with reasonable regard for efficiency? What measures have been taken during planning and implementation to ensure that resources are efficiently used?

- Could the intervention have been implemented with fewer resources without reducing the quality and quantity of the results?

- Could more of the same result have been produced with the same resources?

- Could an altogether different type of intervention have solved the same development problem but at a lower cost?

- Was the intervention economically worthwhile, given possible alternative uses of the available resources? Should the resources allocated to the intervention have been used for another, more worthwhile, purpose?

Chapter 3

This chapter starts with the observation that difficulties in evaluating development initiatives against the goal of poverty reduction in some cases are due to a lack of clarity regarding poverty and poverty reduction rather than insufficient understanding of evaluation. The following topics are briefly discussed from an evaluation point of view:

- The multidimensionality of poverty
- The diversity of poverty within and between societies
- Direct and indirect intervention strategies
- The alignment of development assistance with partner country development strategies and administrative frameworks
- Empowerment and popular participation
- Gender mainstreaming
- Poverty reduction and the environment

Evaluating Poverty Reduction

3.1 Evaluating poverty reduction

For Sweden poverty reduction is the overall goal of development co-operation. Support for human rights and democracy, economic growth, gender equality, health and education, sustainable use of natural resources and protection of the environment are all regarded as part and parcel of the struggle against poverty.[10]

It follows therefore that poverty reduction is also the main standard against which Swedish development co-operation efforts should be evaluated. To put it very simply, if it turns out that a strategy, programme or project supported by Sweden has not contributed to poverty reduction in some significant way, we ought to conclude that it should not have been selected for support in the first place. Similarly, if we wish to claim that one intervention is superior to another, all things considered, we should be able to put forward a good argument to the effect that it has served, or is likely to serve, the cause of poverty reduction better than the other.

When planning to evaluate a development intervention in a poverty reduction perspective, the first thing to remember is that this is no different from evaluating it in relation to any other objective or goal. All that is said in this manual about evaluation in general applies to evaluations focused on poverty reduction. When we find it difficult to evaluate an intervention in relation to its contribution to poverty reduction, it is usually not because we lack any special evaluation tool or technique, although this may also be the case. More commonly, it is because we have not fully considered what poverty is all about, or because it is unclear how the evaluated intervention is expected to help reduce poverty.

---

[10] Directives for Swedish involvement in international development co-operation are laid down in the Policy for Global Development adopted by the Swedish Parliament on December 16, 2003. For development co-operation the main goal is to create pre-conditions for poor people to improve their living conditions. As emphatically stated by the policy, Swedish support for development should have a rights-based perspective and consistently reflect the experiences and priorities of poor people. The following components are singled out as particularly important:

- Democracy and good governance;
- Respect for human rights;
- Gender equality;
- Sustainable use of natural resources and protection of the environment;
- Economic growth;
- Social development and security;
- Conflict prevention and resolution;
- Global public goods.

3.2 Implications of multidimensionality

As understood in the development community, poverty is multidimensional. It not only involves material deprivation and lack of economic opportunity, but also vulnerability and deprivation with respect to health and education, power and influence, social status and human dignity. In Sida's view, poverty is essentially a combination of lack of power and choice and lack of material resources.[11]

This has far-reaching implications for evaluation. When evaluating a development co-operation intervention in a poverty reduction perspective, we need to look at it in relation to all the main dimensions of poverty:

→ Will it increase the assets of poor people, create better opportunities for poor people to earn a living, or otherwise improve their material standard of living?

→ Will it have a significant impact, positively or negatively, on the health and life chances of poor people?

→ Will it provide poor people with education and increase their access to useful and otherwise valuable information and knowledge?

→ Will it strengthen the rights of poor people and make state organisations more responsive to their needs and interests?

→ Will it empower poor people, individually or collectively? Will it increase their ability to assert their rights in relation to the state and more affluent citizens?

[11] Perspectives on Poverty. Sida. October 2002.

→ Will it make poor people less vulnerable to the adversities of armed conflict, natural and humanitarian disasters, market fluctuations, and other untoward developments?

All these questions should be considered during the initial stages of an evaluation, if only in a tentative fashion. In an evaluation of an educational programme, for example, we could start with the educational impact itself. Since illiteracy and lack of knowledge are part of poverty, improving poor people's access to education is an immediate way of reducing their poverty. But, obviously, providing poor people with access to education may not be the sole contribution to poverty reduction of an educational programme. It could also have numerous indirect effects on poverty, both positive and negative. It could make the beneficiaries of the support more competitive in the labour market, for example, or it could empower them in relation to local authorities, both of which could be more important impacts than the direct educational one. A conceivable negative impact is that the programme might benefit relatively privileged people rather than the disadvantaged, and thereby reinforce rather than reduce existing inequalities.

One would expect likely effects on poverty to be well considered in funding proposals and other documents underlying interventions, but occasionally this is not the case. In some cases, positive impacts in relation to poverty are simply taken for granted. In other cases, poverty reduction is the stated goal of the intervention, although it is not clear how the outputs from the intervention are expected to lead to this result. Questions about possible negative effects are often ignored.

Faced with such problems, the evaluators' first task is to reconstruct the intervention logic in the manner suggested in the second part of this manual. Without a clear idea of what the intervention reasonably could be expected to accomplish, formulating evaluation questions and designing a study is bound to be difficult.

3.3 The poor are not a homogenous category

To be poor is always to be deprived of goods that are essential to human well-being. What this actually means, however, may vary significantly from one period, place, society, group, or person to another. As stressed in policy documents and guidelines, an intervention that is intended to contribute to poverty reduction should be based on a careful analysis of the characteristics of poverty in the society, area, or location where it is implemented. Such an analysis revolves around a few key questions:

➜ Who are the poor? What are their characteristics in terms of gender, age, household status, ethnicity, religion, occupation, and so forth?

➜ How are they poor? What is the situation of different groups of poor people in terms of income and consumption, human rights, political power, social discrimination, gender status, health and education, natural resources dependency, occupation, etc.?

➜ What are the obstacles preventing different categories of poor people from moving out of poverty? Lack of economic growth in the surrounding society? Lack of secure property rights with regard to land and other natural resources? Lack of marketable skills and resources? Lack of security? Lack of political power and voice? Discrimination by gender or ethnic origin? Etc.

➜ What is there to build on? How could strategies for coping with poverty be strengthened? What are the resources and opportunities of the poor?

Raised for the first time during planning, questions such as these recur at the stage of evaluation. As evaluators or users of evaluations, we need to know if the intervention to be evaluated is built on a correct understanding of poverty in its local context. Without such knowledge there is no way of telling whether the intervention is relevant in terms of targeting, or if the support that it offers is well designed, given the aspirations and overall situation of the targeted group or category. Without a proper understanding of the local meaning of poverty and the opportunities and constraints facing the poor, a donor may easily end up targeting the wrong categories, or supporting a correctly targeted category in the wrong way.

Note that poverty varies within as well as between societies, and that this is easily overlooked. Development organisations are not immune to overly simplified notions about the characteristics of poverty. In the field of rural development, for example, stereotyped ideas of rural society as an aggregate of small-holder peasant households have sometimes stood in the way of a full appreciation of the complexities of social structures and livelihood patterns in the countryside. A development strategy that is based on such notions is obviously not going to be of much direct help to all those who are landless and depend on other sources of income.

Box 12

Evaluation and HIV/AIDS

When preparing evaluations of projects and programmes in societies severely affected by HIV/AIDS, the epidemic should always be considered. Not only are there important questions concerning the impact of development interventions on the epidemic and the problems that it creates, but there are also important questions about how the epidemic affects the interventions themselves.

Note also that poverty changes and fluctuates over time, sometimes in unpredictable ways, as when a community or society is thrown into a violent conflict or is hit by a natural catastrophe or a major epidemic. The impact of HIV/AIDS in Sub-Saharan Africa is the prime example, creating a development crisis of unprecedented scale. For the international development community as well as for national governments it is a challenge that requires a reassessment of development strategies at every level. The role of evaluation is to provide feedback to those who are responsible for this rethinking and help disseminate useful experience.

3.4 Direct and indirect intervention strategies

We should also consider the fact that strategies for combating poverty may be more or less direct and more or less inclusive. Sida makes a threefold categorisation of poverty reduction interventions:[12]

■ Interventions directly and predominantly focused on different categories of poor people. Examples are targeted safety nets, labour-intensive works programmes, support to refugees and internally displaced persons, and, in some cases, support to non-governmental organisations.

■ Inclusive actions where poor people are affected as members of a broader group of beneficiaries, including sector-wide approaches that are geared to sectors of importance to the poor (education, rural development, small-scale enterprises).

■ General structural approaches aimed at underpinning pro-poor policies. These give support to efforts for democracy and good governance, macro-economic stability, increased accountability, transparency and the fight against corruption.

These types tend to prompt somewhat different kinds of evaluation issues. Where direct interventions and interventions of the inclusive type are concerned, questions of *targeting* are likely to be prominent. If programme benefits are earmarked for a particular category of poor people, we should make sure that all subgroups of the targeted category have benefited, and that people who are not entitled to them have not captured too large a portion of the benefits. In the first case, we are concerned with the issue of coverage – are benefits evenly spread across the target group? Have some parts of the target group been favoured over others? In the second case, we are concerned with leakage of benefits to people outside the target group. Have benefits been appropriated by groups that were not entitled to them?

We must also consider the economic and political costs of targeting. On the economic side, there is an important question whether a targeted programme is a more cost-effective way of reaching programme goals than a non-targeted programme. On the social and political side we have to consider the possible adverse effects of singling out a particular low-status group for support, thereby perhaps antagonising large segments of the surrounding society. As one would expect, sustainability can be a serious problem in programmes that are directly targeted on the poor and provide no benefits to other citizens.

[12] Perspectives on Poverty. pp 9–10.

When dealing with interventions of the inclusive type, on the other hand, it is important to make sure that poor people have equal access to the benefits, and that the programme has not become a means for privileged groups to further strengthen their advantage. While programmes of the inclusive kind are often more likely to be sustainable than programmes intended to benefit a particular group of poor people, they may have their own characteristic weaknesses. Exclusion of poor people is often likely to be one of them. When evaluating programmes of this type it is important to make sure that they have a pro-poor profile and that the cards are not stacked against the poor.

A common problem with regard to interventions of the indirect structural type is that it may be difficult to distinguish those that may have a major poverty reducing effect from those that have only marginal effects on poverty or no such effects at all.[13] It is not unusual that the claim that an intervention will benefit the poor rests on uncertain assumptions about developments in a distant future. When this is the case, evaluators need to be on their guard. The longer the expected causal chain, and the larger the number of problematic assumptions concerning developments outside program control, the more uncertain the impact. When dealing with such complex cases the evaluators' first task is to revisit the intervention logic, and carefully assess the risk analysis against existing expert and local knowledge. Are all the assumptions in the intervention model valid? Is the list complete?

It is important that we understand what can and cannot be accomplished through such an analysis. Assessing the quality of the intervention logic itself is always possible. Is it based on clear and coherent assumptions about contextual factors? Are posited causal links well supported by research and experience? Is the risk analysis reasonable? However, verifying that intended results have been produced as expected is a further and often much more difficult step. Not

infrequently, the conclusion of the analysis of the intervention logic is simply that some or all of our questions cannot be empirically investigated.

As noted in Chapter 2, the difficulties of evaluating results tend to increase as we approach the upper reaches of the hierarchy of objectives. In logframe terminology, it is when the links connecting outcomes to impacts are investigated that the real difficulties appear. This is true for all types of intervention. While the posited links from outputs to outcomes can often be verified – at least if the outcome objectives have been clearly defined in behavioural terms – the connections between outcomes and impacts are often well beyond the reach of evaluation research.

One of the problems is that top-level goals are usually set very high. Project planning manuals recommend that intervention goals should not be set higher than necessary in order to justify the intervention. If we believe that improving communications for an isolated community is good enough as a justification of a road building project this should be the main goal of the intervention. If the construction of the road is justified by an expected increase in trade, that should be the goal.

In development co-operation, however, interventions often have to be doubly justified. Providing a valuable benefit for a particular target group is not enough. We should also be able to explain how the supported intervention is expected to contribute to the achievement of development goals at a national or regional level. Sometimes we are expected to do this even when the intervention is relatively small. How does this particular project in support of a human rights advocacy group contribute to the development of democracy and human rights in the country as a whole? How important is this particular intervention, given alternative strategies for reaching the same overall development goal?

Not much reflection is needed to realise that questions like these are very different from questions about outcomes in relation to more narrowly defined objectives. Clearly, the two types of

---

[13] David Booth et al. Working with Poverty Reduction in Sida. Ms. Stockholm, 2001.

questions may require different types of answers. When faced with questions about development impact at the highest level, the appropriate answer is sometimes that the question itself is ill-conceived. At that level, reassessing the intervention logic may be the best that we can do.

3.5 Integrating donor support with partner country strategies

A further point is that there is a growing consensus in the development community that development co-operation results will improve if the external support from donors is aligned with the poverty reduction strategies and development policies of developing countries. The constraints and opportunities facing donors differ from country to country, of course, and there are countries where the government does not have the capacity to direct the development process as required by the current development agenda. Still, the ambition is the same every-where. The partner country government and other organisations in the partner country should increasingly direct and control the development process; programmes should increasingly replace projects as the dominant mode of aid delivery; and among donors co-ordination should replace fragmentation.

Again, there are clear implications for evaluation. The most important is simply that the efforts of donors to support partner country ownership should be systematically assessed. A second is that there must be evaluations of the larger development co-operation system as well as evaluations of programmes and projects. A third is that questions about national ownership, capacity development, donor co-ordination, integration of external support with national systems of administration and so on should be systematically addressed in project and programme evaluations. Evaluating intended intervention outcomes and impacts is not enough. Interventions should also be assessed in relation to their systems level effects and the agenda for supporting ownership.

Box 13 contains an extensive list of possible criteria for assessing the contributions of donors to the emerging development co-operation agenda. Designed for the evaluation of agency performance, several of these criteria can easily be translated into criteria for the assessment of individual programmes and projects.

3.6 Participation and empowerment

Chapter 1 notes that promoting participatory forms of monitoring and evaluation in programmes and projects supported by Sida is part of a wider Swedish policy of supporting popular participation in development. Here we make the related but different point that the same policy requires that participation be used as a criterion or principle of assessment in evaluations of development interventions supported by Sida.

Box 13

Assessing development agency poverty reduction efforts

Is the development agency's country strategy based on the partner country's own assessment and strategy for addressing poverty? Is it based on a multidimensional poverty concept?

To what extent have the agency's activities been carried out jointly or in co-ordination with other development agencies, bilateral or multilateral?

To what extent have agency administrative and financial requirements been adjusted to, or harmonised with the partner country's existing procedures or with those of other external partners, where these procedures are deemed appropriate?

To what extent has the agency implemented its support in a manner which respects and fosters partner country ownership?

Has the agency supported and strengthened country-led planning, implementation and co-ordination processes?

Has the agency helped facilitate civil society's participation (at local, national, or international level) in debating and deciding the contents of the country's poverty reduction strategy in ways that respect government efforts and concerns?

Has there been a clear, serious commitment of resources to poverty reduction?

Has a commitment been made to provide predictable resources over a medium-term planning timeframe?

Has sufficient care been taken to avoid duplication effort and to build on complementarities across the external development community?

Have efforts been made to improve policy coherence within the agency, and, more broadly, across the full range of DAC Member government ministries and department, and has progress been achieved?

Source: The OECD/DAC Guidelines for Poverty Reduction. Paris. 2001, p. 59.

The stated rationale for supporting popular participation is twofold: participation is an end in itself and participation increases the relevance and effectiveness of development co-operation initiatives. Promoting participation as an end itself is obviously an important part of the overall effort to reduce poverty. While poverty is characterised by exclusion and disempowerment, participation is the very opposite. Support for the participation of poor people is the same as support for their empowerment and inclusion in society on equal terms.

Thus, when assessing development interventions against principles of popular participation we are concerned not only with the ability of poor people to take part in the planning, implementation and monitoring of the interventions themselves, although this is often important. A bigger question is how development interventions affect the ability of poor people to give voice to their interests and participate on an equal basis in society. As suggested in section 3.2 above, this is one of the questions that should always be reviewed when an evaluation is being prepared. Although some interventions are likely to have a much greater impact on structures of governance and participation than others, questions of popular participation are rarely irrelevant.

Note that the issue of participation overlaps with that of ownership. When we talk about national ownership, as distinct from government ownership, we refer to the relationship between the developing country government and its citizens as well as to the patterns of interaction between the former and its external partners. Are public institutions and government policy-

making processes responsive to the interests of poor people and other citizens? Can poor people express their grievances and hold public authorities to account? Without popular participation, national ownership does not exist.

3.7 Gender mainstreaming

Sida has a policy of mainstreaming gender equality in all its work. In the context of evaluation this means two things:

■ Evaluations initiated by Sida should produce information that is relevant to the implementation and follow-up of Sida's policy for promoting gender equality.

■ Evaluations should be designed, implemented and reported in a manner that is consistent with a policy of gender mainstreaming.

A first elaboration of these points is simply that women and men face different types of constraints and opportunities and that this must be reflected in evaluations. When we talk about poor people without further qualification, as we tend to do, gender-based differences are easily overlooked. To avoid that this happens, standard questions concerning relevance, impact, effectiveness, etc. should normally be further defined in gender terms. Thus, instead of asking if a project is relevant to the needs and priorities of the poor, we should ask whether it is relevant to the needs and priorities of both poor men and poor women. In some cases such a specification may not be necessary, but often it will be all the more important.

We should also bear in mind that gender equality is a relational phenomenon involving men as well as women. It will always be important to know if and how the conditions of women (or men) have improved as a result of a particular intervention. Yet, what we need to know as far as gender equality is concerned is how relationships between men and women have changed as a result of the intervention. This point may seem obvious but it is easily ignored. It is not unusual that evaluations fail to consider how recorded changes in the conditions of women (or men) may affect established patterns of gender-based inequality.

A third point is that the pursuit of gender equality is part of the struggle against poverty, and should be evaluated as such. Thus, evaluators must not only examine interventions in relation to what they intend to achieve with regard to gender equality, but should also look at them in relation to the larger goal of poverty reduction. In many cases the link between increased gender equality and poverty reduction may seem obvious. In other cases, however, it may be less apparent. Thus, everything that was said about evaluation of interventions in a perspective of poverty reduction in sections 3.2–3.6 applies in this case as well.

A fourth point concerns the implications of Sida's mainstreaming policy for the planning, implementation and reporting of evaluations. As shown in Box 14 there are important questions about gender inequalities to be considered at all stages of the evaluation process. The practical implications of gender mainstreaming vary with the nature of the evaluated activities and the purpose of the evaluation itself. If mainstreaming turns out to be time-consuming or otherwise demanding, as it may, this simply shows that gender issues are of great importance to the evaluation.

Notice, finally, that gender mainstreaming can itself be made the subject of evaluation. If an organisation has adopted a policy of gender mainstreaming, it goes without saying that it may also want to evaluate its implementation.

Box 14

Gender issues in evaluation

PREPARATION

What does the funding proposal and related documents say about the importance of the intervention with regard to poverty reduction and gender equality? What is the project expected to achieve in these terms?

According to the same sources, what is the importance of gender-related factors in the implementation of the intervention?

What significance should the evaluation assign to questions about gender equality, given what we know about the intervention and its purpose?

What is the evaluability of identified gender equality issues? Are there baseline data for assessing gender equality impact?

Do the ToR clearly specify what the analysis of gender should include?

Is the recruited evaluation team competent to assess the gender equality issues in the ToR? Is their approach to gender sufficiently described in the tender documents?

THE RESEARCH PHASE

Do the evaluators consider the fact that men and women often communicate differently, and that women may not be able to express themselves freely in all situations?

Are gender-sensitive methods of data collection used where required?

Do the evaluators consult with both men and women?

Are data properly disaggregated by gender?

REPORTING

Does the report answer the questions about gender equality in the ToR?

Does the analysis of gender issues match the importance of these issues in the ToR?

Are findings about gender equality assessed in relation to the standard evaluation criteria of effectiveness, impact, relevance, sustainability, and efficiency?

FOLLOW-UP, DISSEMINATION AND USE

Has the need for gender-sensitive methods of feedback and communication been considered?

Has the evaluation report been disseminated to all interested parties?

Are conclusions and lessons concerning gender equality fed back into operations and spread to others who may be interested in them?

3.8 The environmental perspective

Sida has a policy of integrating a concern for the environment in all its activities, and it also supports development interventions that specifically promote a sustainable use of natural resources and protection of the environment.

As explained in Sida's Guidelines for the Review of Environmental Impact Assessments,[14] all Sida contributions should include an environmental impact assessment (EIA). Made at the stage of planning and appraisal, such an assessment is intended to make sure that benefits to poor people do not entail unacceptable costs in terms of environmental degradation, now or in the future. It also looks for positive linkages between a policy of environmental sustainability and poverty reduction. Measures to improve the environment can lead to poverty reduction, just as poverty reduction efforts can have positive effects on the environment.

The questions about environmental impact raised during planning and appraisal should be revisited when evaluations are initiated. The original prospective assessment should not be taken for granted. It may have been invalidated by unexpected developments during the implementation phase or it may have been incorrect all along. The main question is how the evaluated intervention affects existing linkages between poverty and the environment. In some cases, certain infrastructure projects, for example, the impact may be direct and easy to identify, in other cases, such as a sector reform programme or a health care project, it may be more difficult to detect, although no less important.

> "The main question is how the evaluated intervention affects existing linkages between poverty and the environment"

The scope of the assessment varies greatly with the type of intervention being evaluated. In some cases, the environmental issues occupy the centre stage of the evaluation, in other cases they can be summarily dealt with in the preparatory phase. Yet, in no case can they be entirely disregarded. Box 15 lists questions about environmental consequences and their management as they appear at different stages of the evaluation process. Further guidance can be found in Sida's Guidelines for the Review of Environmental Impact Assessments.

[14] www.sida.se/publications

Box 15

Evaluating environmental consequences

PREPARATION

What was the conclusion of the ex-ante Environmental Impact Assessment (EIA)? Has a follow-up of environmental consequences been made?

Should environmental issues be assigned a central or marginal position in the evaluation, given what we know about the potential environmental consequences of the intervention and their relation to poverty reduction?

Does the EIA provide sufficient baseline data to make a satisfactory evaluation of environmental impact possible? How can we compensate for weaknesses in the EIA?

Do the ToR contain clear directives for the evaluation of environmental issues?

Do the ToR make clear that the environmental analysis should cover the actual and potential impact of the intervention on the conditions of the poor.

Do the tenders from interested evaluators reflect a proper understanding of the environmental issues at stake in the evaluation? Does the evaluation team have the competence and experience required for an assessment of the intervention in the perspective outlined in the ToR?

Have the evaluators been properly briefed about Sida's Guidelines for the Review of Environmental Impact Assessments and other relevant instructions regarding the environmental dimensions of development interventions supported by Sida?

THE RESEARCH PHASE

Have Sida and its partners provided the evaluators with the necessary support to evaluate the environmental issues as required by the ToR?

REPORTING

Does the report adequately answer the questions about environmental consequences in the ToR? Have both positive and negative consequences been considered?

Does the report answer questions about the analysis and management of environmental issues during the planning and implementation of the intervention?

Does the report contain a reassessment of the EIA carried out during planning?

Are findings regarding environmental dynamics assessed in relation to the standard evaluation criteria of effectiveness, impact, relevance, sustainability, and efficiency?

FOLLOW-UP, DISSEMINATION AND USE

What are the practical implications of findings and conclusions regarding environmental impact and the management of environmental issues by Sida and its partners? What are the lessons learned?

Are findings and conclusions regarding environmental issues properly reflected in Sida's management response to the evaluation?

Are the findings and conclusions of the evaluation properly disseminated within Sida and effectively made available to Sida's partners?

Part Two:
The Evaluation Process
Step by Step

The Evaluation Process Step by Step

The step-by-step guidelines presented in this part of the manual are designed to help Sida and its partners manage evaluations. The guidelines cover the main steps of the evaluation process, and provide practical advice on how evaluations can be tailored to the needs and interests of their intended users.

Step 1	Step 2	Step 3	Step 4	Step 5
Initial considerations	Preparation of evaluation tasks	The evaluation research phase	Reporting and dissemination	Management response
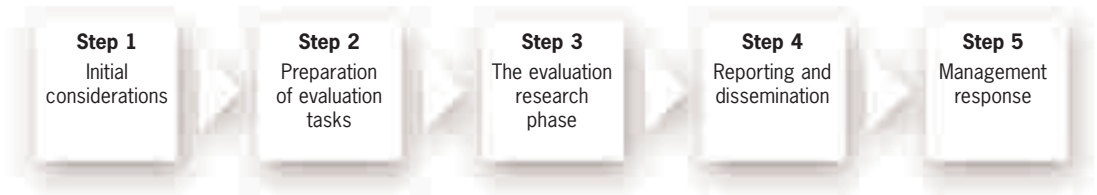
The guidelines are divided into five sections corresponding to the five main steps of the evaluation process. Each section begins with an overview of the main tasks during that particular step. The tasks are then further described in a series of sub-sections. At the end of each sub-section, there is a checklist that briefly summarises the tasks from a practical "how-to-do-it" perspective.

The guidelines are primarily directed to *evaluation managers*, i.e. Sida staff and others who are responsible for managing and co-ordinating evaluation initiatives. In many cases the evaluation manager is identical with the programme officer managing Sida's support to the evaluated intervention. As we are focusing on roles and relationships in the evaluation process, however, we consistently refer to the person administrating the evaluation as the evaluation manager.

The guidelines are relevant for most types of evaluation. The described tasks are essentially the same, irrespective of whether the object of the evaluation is a project, programme, intervention theme, or aid modality.

Initial Considerations

In this first step of the evaluation process, the most important tasks of the evaluation manager are to:

1.1 Involve interested stakeholders,

1.2 Define the purpose of the evaluation,

1.3 Establish an organisation for evaluation management.

Step 1 Initial considerations	**Step 2** Preparation of evaluation tasks	**Step 3** The evaluation research phase	**Step 4** Reporting and dissemination	**Step 5** Management response

1.1

Involving stakeholders

As briefly discussed in Chapter 1.7, evaluations commissioned by Sida should be carried out in a spirit of partnership. Sida's co-operation partners and other important stakeholder groups should be encouraged to actively participate in the evaluations that Sida initiates. For Sida, stakeholder participation is an end in itself as well as a means.

The pragmatic argument for participation is that the quality of evaluations tends to improve when co-operation partners and other stakeholder groups are actively involved in the evaluation process. Among the benefits of participation are greater accuracy and depth of information, increased credibility and acceptance of findings, and better correspondence to the practical concerns of stakeholders.

Stakeholder participation may, however, increase both cost and time. Therefore, the evaluation manager and others responsible for the evaluation should discuss to what extent different stakeholder groups should participate, given their legitimate interest in the evaluation as well as costs, timing and other practical aspects.

Stakeholder groups in evaluations

CO-OPERATION PARTNERS

The parties that request donor support and that are responsible for planning, implementing and following up the evaluated intervention.

PRIMARY STAKEHOLDERS

Target groups who benefit from the results of the evaluated intervention (beneficiaries), but also those groups of people who may be adversely affected.

DONOR ORGANISATIONS

Sida and other donor organisations that provide financial, technical and other types of support to the evaluated intervention.

INTERESTED PARTIES

Groups that have other stakes in the evaluation, for example partner governments, implementing consultants, and organisations channelling donor support.

Involving co-operation partners

When Sida initiates an evaluation, and starts to involve interested stakeholders, among the first steps are consulting with co-operation partners. Through these consultations Sida, firstly, establishes what interest co-operation partners have in the evaluation. Do they wish to be actively involved in the evaluation process, and do they want the evaluation to be geared towards their own management purposes? If yes, a joint evaluation of the kind discussed in Chapter 1 could be a suitable arrangement for evaluation management. Secondly, Sida and its co-operation partners may have to discuss how other partner country stakeholders should be involved in the evaluation process.

Involving co-financing donors

When Sida is one of several development agencies supporting the intervention, it should, as far as possible, seek to co-ordinate its evaluations with those of the other agencies. To avoid unnecessary duplication of efforts, the evaluation manager should examine whether the planned evaluation can be jointly carried out or if there are other means of collaboration. In some cases, it may be possible for Sida to use information already produced by other agencies, rather than to undertake a study of its own. As a rule, however, co-operation in evaluation should be discussed well in advance of any particular study.

Involving primary stakeholders

It is important to consider how target groups and other primary stakeholders should be involved in the evaluation. This should be done as early as possible, even when target groups cannot realistically be expected to actively participate in preparatory

work. It is essential that the evaluation is designed to enable target groups to constructively participate in the research process and express their point of view. Note that target groups are often neglected in evaluation processes, even when they represent a key source of knowledge and information.

Involving other interested parties

Evaluations may have impacts also on stakeholder groups that do not participate actively in the evaluation process and do not belong to the intended beneficiaries. Evaluation ethics require that such groups are informed about the evaluation and given real opportunities to express their interests.

Using stakeholder resources efficiently

Every evaluation is dependent on the co-operation of stakeholders for information and practical arrangements. For example, co-operation partners and implementing consultants are often requested to provide documentation and prepare meetings. They are almost always expected to be interviewed and share their views on intervention performance and results.

The evaluation manager must ensure that the evaluation does not overtax stakeholders' time and resources. Stakeholders should be informed as early as possible about the evaluation. This will assist them in planning their inputs to the study without straining other commitments and duties.

CHECKLIST

for involving key stakeholders

☐ Carry out a preliminary stakeholder analysis, and discuss with partners as early as possible how key stakeholder groups might wish to be involved in the process.

☐ Inform major stakeholder groups about the evaluation initiative without unnecessary delay.

☐ Indicate, clearly and as early as possible, how different stakeholder groups are expected to contribute to the evaluation with information and practical support.

1.2 Defining the evaluation purpose

Defining the evaluation purpose is one of the most important tasks in the early stages of the evaluation process. A clear purpose helps the formulation of evaluation questions, and makes it easier for external evaluators to produce a study that Sida and its partners may find relevant and useful.

The intended use of the evaluation

The evaluation purpose should be formulated in a way that specifies how the information produced by the evaluation is to be used. "To analyse project effectiveness and impacts", "assess the relevance of the project", and the like are descriptions of the means by which evaluation purposes can be achieved; they are not descriptions of the purposes themselves. The purpose of an evaluation is always an action that the evaluation is intended to make possible.

As noted in Chapter 1 the purpose of development cooperation evaluations can be summarised in terms of accountability and learning. Below are some concrete examples of purposes for which Sida initiates evaluations.

Examples of Sida evaluation purposes:

- Provide Sida and its partners with an input to upcoming discussions concerning the preparation of a second phase of the evaluated intervention.

- Help Sida and its partners make sure that the intervention is well on track and is likely to reach its objectives.

- Help Sida decide whether support to the intervention shall continue or not.

- Provide Sida with relevant and useful background information for an annual sector review.

- Gather data about the effectiveness and impacts of the evaluated interventions in order to help Sida elaborate a results-based country strategy.

- Provide Sida and its partners with lessons that can be used in policy work or when designing programmes and projects elsewhere.

- Provide information about results that Sida and its partners can use in reporting to principals and the general public.

If no similar purpose can be identified, it may signal that the evaluation is not sufficiently useful to justify the investment. In such cases, the evaluation could be postponed until it can better be fed into operative and policy-making processes. Perhaps it should not be carried out at all. Clearly, there is no point in carrying out evaluations unless they can be used productively.

Consulting with partners and other possible users

When Sida initiates an evaluation it has a tentative idea of how it intends to use the study. This, however, does not mean that the definition of the evaluation purpose always is Sida's exclusive

responsibility. If Sida's co-operation partner or a co-financing donor wishes to participate actively, the definition of the purpose becomes a joint undertaking. If the information needs of the prospective partners are too diverse to be accommodated in one and the same evaluation, the idea of a joint study should be abandoned. Where accountability is the main purpose of the evaluation, a joint study may also not be the best option.

Several rounds of consultations must often be held before an acceptable definition of the purpose or purposes of the evaluation is found. This process must be allowed to take time. Most Sida evaluations are listed in Sida's annual evaluation plan, often months before details are elaborated. Therefore, Sida usually has enough time to consult with relevant stakeholders.

Note that the timing of the evaluation is likely to be important. The evaluation must be phased in such a way that the participants can make good use of it. If the evaluation cannot be completed before it is to be used, it should be redesigned or cancelled. It must be taken to account that different stakeholders may have different time frames and deadlines.

CHECKLIST

for the definition of the evaluation purpose

☐ Identify possible users of the evaluation, i.e. those groups that are expected to make use of the evaluation process and its results.

☐ Ensure that the evaluation purpose is defined through a participatory process that engages all interested users of the evaluation.

1.3 Establishing an organisation for evaluation management

In many cases, the necessary contacts between the evaluation manager and different stakeholder groups can be maintained without establishing a formal organisation for evaluation management. However, in complex evaluations, where several stakeholder groups have major interests, a reference group or steering group is usually formed.

While a reference group is an advisory body, a steering group is formed to give stakeholders an opportunity to participate in deciding key issues during the evaluation process. Involving representatives of major stakeholder groups is likely to enhance the credibility of the evaluation and may also create a more active interest in the results of the study among those involved.

A typical steering group would include representatives of at least one of the following groups: co-operation partners, implementing organisations and co-financing donors. These are typical tasks for a steering group:

- Providing inputs to the terms of reference,
- Formally accepting the terms of reference,
- Monitoring the recruitment of external evaluators,
- Approving the selection of evaluators,
- Commenting on draft reports and
- Approving the final report.

☐ In consultation with stakeholders establish practical arrangements for communication and co-operation during the evaluation process.

☐ Consider forming a reference group or a steering group when the evaluation is complex, involves several stakeholder groups, and the intended users include non-Sida stakeholder groups.

CHECKLIST

for establishing an organisation for evaluation management
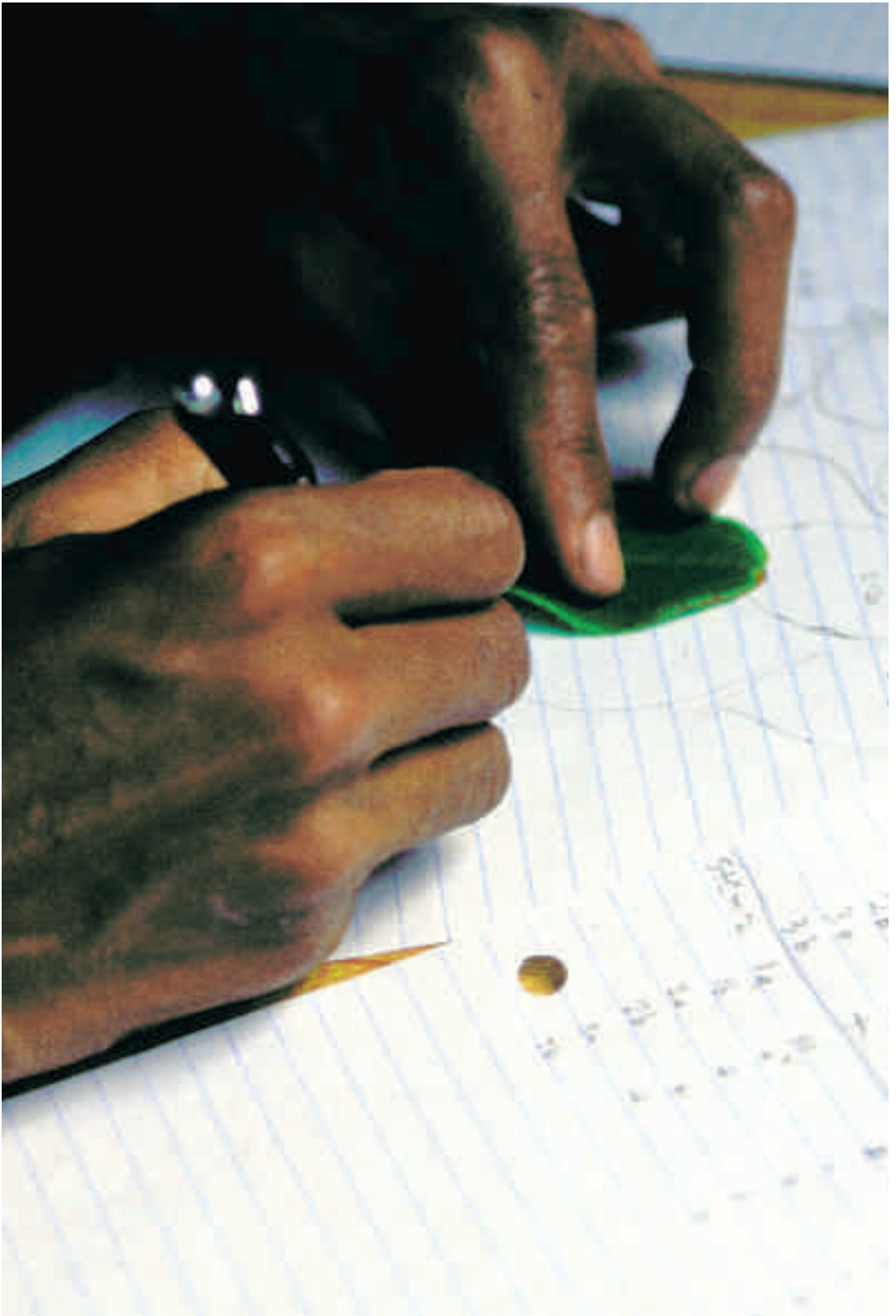
Preparation of Evaluation Tasks

In this second step of the evaluation process, the most important tasks for the evaluation manager are to:

2.1 Review the intervention selected for evaluation,

2.2 Define the questions that the evaluation should answer,

2.3 Assess evaluability, i.e. the extent to which these questions are answerable,

2.4 Consider the option of dealing with the above tasks in a pre-study,

2.5 Estimate the total budget of the assignment,

2.6 Formulate terms of reference for a team of external evaluators and

2.7 Recruit a team of evaluation consultants.

Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**
Initial considerations	Preparation of evaluation tasks	The evaluation research phase	Reporting and dissemination	Management response

The complexity of the preparatory tasks varies from one evaluation to another. For example, if the evaluation manager is already familiar with the intervention and its objectives, and the intervention is a relatively free-standing project of limited scope, the initial review is usually a fairly simple task. Also, if the evaluation is geared towards planning and implementation processes, rather than outcomes and impacts, the assessment of evaluability may not be very complicated.

If the preparatory tasks turn out to be complex and time-consuming, the manual suggests alternative ways of handling them. First, there is always the option of including tasks 2.1–2.3 in a separate pre-evaluation study carried out by consultants. Second, there is the alternative or additional option of including some of these tasks in the inception phase of the evaluation assignment itself. Directions for these two options are given in Sections 2.4 and 3.1 respectively.

Reviewing the intervention

It is important to review, as early as possible, the main features of the intervention selected for evaluation. Without a good understanding of the intervention and its intended logic, we cannot identify which questions that should be answered by the evaluators.

Time periods and intervention phases

Development interventions often consist of distinct but related phases. While a phase may cover a period of 2–3 years, the intervention itself may date back several years and even decades. The question of when individual interventions actually start and end is not always easy to answer. Still, it is important that the terms of reference for the evaluation are precise about what periods and phases the study should deal with. The decision may depend on a variety of factors, not least which information we expect from the evaluation. This must be discussed between the intended users of the evaluation.

The intervention logic

When preparing the evaluation, the evaluation manager should summarise the intervention logic in the conventional logframe format, including planned activities, expected outputs, outcomes and impacts, as well as the indicators and assumptions on which the intervention logic is based. Without a preliminary analysis of goals and objectives, indicators and assumptions, etc, it is difficult to pose the right questions. A preliminary analysis of the intervention logic will also be useful to the evaluators. Otherwise, they may waste plenty of time in figuring out what they actually should be evaluating.

Background information for an analysis of the intervention logic is found in project documents and Sida's rating system (SiRS). The rating system summarises the intervention logic, and describes the outputs of all interventions with a total budget exceeding SEK 3 million, and an agreement period of at least three years.

When project documents and SiRS lack required information, consultations with co-operation partners and other stakeholder groups may help fill the gaps. Co-operation partners and other stakeholders may also be asked to check that the documentary analysis is correct.

Activities and outputs

Information about activities and outputs is necessary when preparing evaluations. It allows us to check if the identified intervention logic indeed is plausible and hence useful for guiding the evaluation. There is little point in carrying out an evaluation on the basis of an intervention logic that does not correspond to what actually took place during implementation.

In many cases, a good way to gather information about activities and outputs is to review the budget and activity statements that partner organisations send to Sida in the form of quarterly, annual and final progress reports. The review should identify, as far as possible, the content, budget and total expenditure of distinct activity and output categories.

Note that SiRS can also be helpful in this respect. However, since it is limited to data on six outputs (also for more complex interventions) and does not automatically provide information about activities and the budgets and disbursements per each output, it must be supplemented by an analysis of progress reports.

Complex interventions

Interventions supported by Sida often embrace several activity and output categories, or components, each with its own distinct set of objectives. To the extent that these components have a common objective, such interventions are normally referred to as programmes.

Complex interventions such as these represent a special challenge. If all the components of a complex intervention were to be evaluated, the evaluation process could itself become very difficult, and perhaps unacceptably time consuming and costly. It may therefore be necessary to focus the evaluation on a limited number of components. In many cases, the evaluation purpose can be achieved if the study is limited to components that are financially important or central to the main theme of the intervention.

☐ Specify what intervention phases the evaluation should deal with, and make sure that the intervention logic, and the activities and outputs during those phases are reviewed in detail.

☐ Review planning documents and SiRS data and summarise the intervention logic with the normal distinction between planned activities, outputs and effects.

☐ Review SiRS data and co-operation partners' progress reports and summarise the content, budget and expenditure of the activities and outputs implemented and produced to date.

☐ Analyse the extent to which the activities carried out can reasonably be expected to meet the objectives of the intervention.

☐ If necessary, consult with co-operation partners to obtain more specific, plausible and up-to-date accounts of intervention objectives and assumptions.

☐ Consider focusing on key components when the intervention is complex and involves several components.

☐ If these tasks are likely to be complex and time-consuming, consider commissioning a pre-evaluation study from external consultants.

Formulating evaluation questions

2.2

An evaluation should not seek to answer more questions than required for its practical purpose. Limiting the focus and scope of the study helps ensure an efficient use of evaluation resources. On the other hand, a narrow focus or scope may not be fully compatible with a participatory approach. Where several stakeholder groups are involved, the list of questions to be answered by the evaluation is likely to become longer.

Regardless of the number of stakeholders, however, it is necessary to formulate evaluation questions and choose what evaluation criteria, policy issues, and standards of performance should be included in the evaluation. The key question for this step in the process can be formulated as follows: Which information is required in order to satisfy the practical purpose for which the evaluation is undertaken?

Evaluation criteria

The criteria of relevance, effectiveness, impact, sustainability and efficiency discussed in Chapter 2 provide us with a useful point of departure. These criteria are basic yardsticks for evaluation and can normally be accepted as such by all stakeholders. Each one represents an important general question that is relevant to every development intervention:

EFFECTIVENESS

Has the intervention achieved its objectives or will it do so in the future?

IMPACT

What are the overall effects of the intervention, intended and unintended, long term and short term, positive and negative?

RELEVANCE

Is the intervention consistent with the needs and priorities of its target group and the policies of the partner country and donor agencies?

SUSTAINABILITY

Will the benefits produced by the intervention be maintained after the cessation of external support?

EFFICIENCY

Can the costs of the intervention be justified by the results?

As stated in Sida's Evaluation Policy, these questions should always be considered when Sida initiates an evaluation.[15] They are not compulsory, but none of them should be set aside without a tentative assessment of their bearing on the evaluation. In many cases there are good reasons to include all five in the study. Chapter 2 discusses them in detail and provides examples of more specific evaluation questions related to them.

Policy issues

The Swedish policy for development co-operation, with its main goal of poverty reduction, is a second major point of reference for the formulation of evaluation questions. Since all interventions supported by Sida should contribute to poverty reduction in one way or another, questions about poverty reduction should be discussed in virtually all evaluations initiated by Sida. If nothing else, the analysis of the intervention logic should indicate how the evaluated activities are intended to serve this end. Normally, one would also expect the evaluation to make an assessment of the plausibility of the intervention logic.

Chapter 3 discusses what evaluating interventions against the poverty reduction goal may entail. It underlines the importance of assessing interventions in relation to a multi-dimensional and rights-based concept of poverty and reviews a number of issues that should always be taken into consideration when poverty reduction is the standard of evaluation. It also notes that support for gender equality and protection of the environment are important elements of the poverty reduction goal, and briefly discusses their implications for evaluation.

[15] www.sida.se/publication

Suggestions from other evaluations

Reviewing evaluations of similar interventions elsewhere or evaluations of other types of intervention in the same society or region can be a useful way of identifying pertinent evaluation questions. It can help us identify design factors that could be important for the success of the intervention, and make us sensitive to aspects of culture and social structure that may affect its implementation and results. Building on lessons from previous evaluation work is essential if the evaluation is intended to contribute to wider discussions about good development practice.

Recommendations and lessons

Given a well-defined purpose it may be self-evident what kinds of recommendations and lessons the evaluation should produce. However, it is generally advisable to be explicit in this respect. Useful lessons are not produced automatically. If we want the evaluators to reflect on the significance of the evaluated intervention in relation to wider concerns of some kind, we must ask them do to so. Note that asking the evaluators to consider a particular issue is not the same as preventing them from raising issues that are not specifically mentioned in the terms of reference. As a rule, the evaluators should be encouraged to make useful contributions of their own.

☐ Consider which information is required in order to satisfy the purpose for which the evaluation is undertaken.

☐ Decide what evaluation criteria and policy issues should be used in assessing the selected intervention.

☐ Review evaluations of similar interventions elsewhere and evaluations of other types of interventions in the same society or region.

☐ Consider which recommendations and lessons are particularly relevant to the evaluation purpose.

CHECKLIST

for the definition of evaluation questions

Assessing evaluability

2.3

The main objective of the evaluability assessment is to find out to what extent the proposed evaluation questions can be answered. This assessment often leads to modifications of the evaluation design. Some evaluation questions will then be withdrawn as impossible, overly difficult or excessively costly. Other questions will have to be further elaborated.

Evaluability is usually a matter of degree; many or most evaluation questions can be answered to some extent. Note also that in many cases definitive answers about evaluability cannot be expected at this early stage of the evaluation process. Often,

questions about evaluability must be further examined during the research phase of the evaluation (Section 3.1). Still, the earlier questions about evaluability are discussed and answered, the better.

Of the many factors that determine evaluability, the following tend to be particularly important: the specificity of intervention logic; the existence and quality of baseline and implementation data; the availability of key informants; and the timing of the evaluation in relation to the intervention cycle.

The specificity of the intervention logic

Most evaluations include an analysis of project effectiveness, i.e. an assessment of whether planned outputs, outcomes and impacts have been realised as expected. As discussed in Chapter 2, such an analysis requires that the objectives of the intervention clearly state what the intervention is expected to achieve in terms of observable developments. In some cases, there are verifiable indicators at every level of the goal hierarchy. If the planning documents do not contain such indicators, clarifications from co-operation partners and other stakeholders, are necessary as suggested in Section 2.1.

The availability of relevant data

The availability of relevant information is a key issue in every evaluation. Baseline information is necessary for outcome and impact-oriented evaluations, as they provide benchmarks against which developments can be identified. Implementation data that clarify intervention activities and outputs are important in all types of evaluation. In studies of impact we may also need data that can be used to estimate the counterfactual. This is further discussed in Chapter 2.

Collecting existing documentation is crucial for evaluation efficiency. Even if the intervention has never been evaluated before, there may still be useful information available, for example internally produced monitoring data or information produced by external monitoring teams. Taking stock of existing information economises with stakeholder resources, as repetitive and overlapping interviews can be avoided.

Access to key informants

Successful evaluation requires that different stakeholders are willing and able to participate in the evaluation. Access to key informants may prove difficult because of, for example, staff changes, vacations, or problems of travel and communication. Stakeholders may also lack interest in participating in the evaluation process.

It is therefore imperative to check if key informants are available before the evaluation assignment starts. If co-operation partners or other stakeholders are unavailable, it may be impossible to implement the study as planned.

The timing of the evaluation in relation to the intervention cycle

This is clearly an important determinant of evaluability. If the intervention is still at the implementation stage some of the expected outcomes and impacts have probably not yet occurred and can therefore not be subject to study. In some cases, the full effects of the intervention will not have occurred until many years after its completion. Evaluations may always assess the likelihood that an expected future impact will occur, of course, but a study of a planned future impact is not the same as a study of real impact.

☐ Assess the extent to which the intervention logic provides evaluators with operational benchmarks against which outputs, outcomes, impacts and assumptions can be evaluated.

☐ Check whether necessary baseline and monitoring data is available.

☐ Check the availability of key informants, such as planners, intervention staff, and target group representatives.

☐ Assess the extent to which the evaluation questions can be answered, given the timing of the evaluation in relation to the current phase of the intervention cycle.

☐ Keep in mind that it may not be possible to make a full evaluability assessment at this early stage. Consider the possibility of completing the evaluability assessment during the research phase.

CHECKLIST

for the evaluability assessment

Making a pre-evaluation study

2.4

Reviewing the intervention, formulating evaluation questions, and assessing evaluability are important tasks, all of which should be carried out before the terms of reference are finalised. In many cases, however, they require considerable efforts. Where the intervention is complex or where the scope and focus of the study must be negotiated with several different stakeholders, the possibility of commissioning a pre-evaluation study from an external consultant should be considered.

A pre-evaluation study can include representatives of Sida and the partner organisation responsible for the intervention. Assisted by an external consultant, Sida's evaluation manager and a representative of the co-operating organisation can jointly review the intervention and attempt to identify questions that

both parties find useful. In many cases, this might be a more constructive way of promoting partner country participation in Sida evaluations than to solicit inputs to terms of reference already drafted by Sida.

While several of the tasks described in Sections 2.1–2.3 can be carried out with the assistance of an external consultant, major decisions regarding the scope and focus of the study rest with Sida and its partners in the evaluation. Note that not even a pre-evaluation study may be able to fully assess the evaluability of all the preliminary questions. In some cases, questions of evaluability cannot be answered before the evaluation has started.

CHECKLIST

for a pre-evaluation study

- ☐ Consider if there are any major preparatory tasks that cannot be carried out without an externally commissioned pre-evaluation study.
- ☐ Consider if a pre-evaluation study is compatible with the overall time schedule and budget of the evaluation.
- ☐ If a pre-evaluation study is carried out, remember that key decisions about the evaluation rest with Sida and its partners, and that the pre-evaluation study is only one of several inputs when formulating the terms of reference.

2.5

Estimating evaluation costs

When the evaluation questions have been formulated, the next step is to estimate the costs of the evaluation assignment. Setting a budgetary limit for the consultant's assignment is important for several reasons. First, it helps indicate the level of ambition for tasks associated with the evaluation. Second, it is necessary for Sida's overall financial planning. Third, a budgetary limit is needed for decisions on correct procurement procedures.

Budgetary limits do not seriously undermine cost competition. Even if all bids are close to the given upper limit for the assignment, the bidding consultants still compete in terms of weekly fees and the number of working weeks that they propose to spend on the assignment.

The most important cost item in an evaluation budget is the consultant's fee. The evaluation manager can arrive at a budget limit by estimating the total number of person-weeks suitable for the assignment. By multiplying the number of weeks with a competitive weekly fee, an approximate fee limit can be calculated.

The time required for evaluation assignments is often underestimated. In many cases, the information necessary to answer evaluation questions is not readily available, but has to be collected by the evaluators through interviews with stakeholders and time-consuming analyses of secondary data.

With the relatively few person-weeks used for evaluations commissioned by Sida – typically three to six person-weeks in the field – even a small number of evaluation questions may well stretch the limits of the possible. From a purely technical point of view, it is usually advisable to examine a smaller number of questions in depth rather than a larger number more superficially.

When estimating the time required for the evaluation it is a good idea to consult with more experienced colleagues and external expertise. It should be noted, however, that estimates of the time necessary for an evaluation generally tend to be somewhat arbitrary. When there is much uncertainty about the time needed for the study, the option of a flexible contract should be considered.

Notice that the budget limit shall cover reimbursable costs, such as travel, hotel and other costs, as well as fees. Of the total reimbursable costs, a 10% contingency item should be included in the estimated budget.

☐ Set a budget limit for the evaluation.

☐ Use standards such as a competitive consultancy fee and an estimated amount of person-weeks for the evaluation when setting the budget limit.

CHECKLIST

for estimating evaluation costs

Writing terms of reference

2.6

The terms of reference (ToR) summarise the results so far of the preparatory tasks undertaken. The ToR also outline a work plan and time schedule for the assignment, the required competence and composition of the evaluation team, and the reports and other outputs that the evaluators should deliver under the contract.

The ToR may also indicate what kind of methodology should to be used by the evaluators for data collection and analysis. In most cases, however, it is the responsibility of the evaluators to elaborate a suitable methodology and work plan, either in the tender document or in an inception report.

The importance of investing sufficient resources in the preparatory steps of the evaluation, and of documenting the results in the ToR, cannot be overemphasised. The ToR constitute the evaluation manager's main instrument in instructing evaluators how the assignment should be carried out. They also serve to document the main points of the agreement between Sida and its partners in the evaluation.

Formulating the ToR must be allowed necessary time, and be based on consultations with concerned stakeholder groups. If a reference group has been formed, its comments on the ToR may be the group's first major input to the evaluation. If a steering group is involved in the direction of the evaluation, the stakeholder representatives must have real opportunities to influence the ToR. Normally, several rounds of consultation and redesign are needed before the ToR can be finalised.

The ToR must be formulated in a way that facilitates mutual understanding between the evaluation manager and the evaluators on all matters of importance. Ambiguous ToR, resulting in misunderstandings between the evaluation manager and the consultant, are a common problem that tends to be difficult to solve if not deliberately tackled in the preparatory phase.

Before the ToR are finalised, we must examine them from the point of view of prospective bidders. The following are some of the questions that evaluators are likely to regard as important:

- Can the evaluation questions be answered? Are there any provisions for the possibility that some of them may not be answerable? Are the ToR flexible enough to allow for contingencies of this kind?

- Are there any non-negotiable directives in the ToR regarding evaluation methods or other aspects of the evaluation process? Are they reasonable?

- Which expert knowledge is required for the job?

- Have sufficient time and resources been allocated for the evaluation?

- Which practical support will the evaluators receive from the client and other stakeholders during the evaluation process?

- Are client expectations regarding reporting clearly formulated?

- What do the ToR imply with regard to the interaction between the evaluators and the client? Will the evaluators have a substantive role in the design of the study? Are they expected to assess the intervention from a truly independent point of view? Is their mandate more limited?

If the ToR have been well formulated, prospective bidders will be able to find answers to all of these questions.

☐ Consider the guidelines for the writing of ToR found in Annex A of this manual, as well as on Sida's Intranet (Sida Templates – avdelnings-specifika mallar/UTV).

☐ Use Part 1 of this handbook for explanations of evaluation concepts and as a general source of ideas for the formulation of the ToR.

☐ Make sure that the ToR provide all the information necessary for prospective evaluators to decide whether the assignment matches their interests and qualifications.

Recruiting evaluators

2.7

Sida typically contracts out its evaluations. Evaluators are selected through a procurement process that is intended to identify candidates suited to deliver expected results. The process is subject to the same rules of procurement as other Swedish development co-operation activities.

The selection of the evaluators is an important step. If evaluators are not qualified for the assignment, work in involving stakeholders and formulating the ToR will be wasted. If the evaluators are well qualified, on the other hand, we may receive good evaluations even with vaguely formulated ToR.

The skills and other qualifications needed by the evaluators vary from case to case, but the following are usually important:

EVALUATION EXPERTISE

This is a package of skills. It includes the conceptual and methodological skills required for successful evaluation research.
It includes the communicative skills necessary for creating rapport with stakeholders, facilitating stakeholder participation, and effectively presenting evaluation results to diverse audiences. It also includes the organisational skills necessary for planning and managing an often complex evaluation research process involving many people.

SUBJECT MATTER EXPERTISE

This is always important, although more so in some evaluations than in others. For example, in an evaluation of an intervention to encourage police officers to alter attitude and behaviour towards the poor, some knowledge of factors determining police attitudes and behaviour would be required. It is not until the evaluation questions have been formulated, however, that the need for subject matter expertise can be more precisely defined.

LOCAL KNOWLEDGE

Since determinants of project success are often local in nature, a good understanding of local social and cultural conditions can be very important. For example, in an evaluation of efforts to reform police behaviour in South Africa, some knowledge about apartheid and South African politics of ethnicity would no doubt be useful. Normally, evaluators also need local knowledge to be able to successfully interact with stakeholders. When the evaluation involves contacts with street-level officials or representatives of target groups, language skills may be required.

INDEPENDENCE AND DETACHMENT

Independence from the object of evaluation and freedom from bias are important requirements regardless of the purpose of the evaluation. Along with the skills mentioned above, both are determinants of the credibility of the evaluation. If the evaluation team is thought to be unqualified, its members regarded as culturally or ideologically biased or too closely involved with the client or other stakeholders, credibility will suffer.

Since the CVs attached to the proposals from bidding evaluators may not contain all the information needed for identifying the evaluation team that is best suited for the job, the tender documents should include a note that explains how the skills and experiences of the team match the requirements listed above.

The evaluation team should be recruited well before the evaluation takes place. The supply of evaluators is limited and well qualified evaluators must often be contracted well in advance. Recruiting evaluators at the last minute may considerably reduce prospects for obtaining a good evaluation.

CHECKLIST

for recruiting consultants

- ☐ Formulate required skills and qualifications of evaluators in the ToR.
- ☐ Make sure that bidders have the required skills and that they satisfy the requirements of independence and detachment.

The Evaluation Research Phase

In this third step of the evaluation process, the most important tasks of the evaluation manager are to:

3.1 Supervise the inception period of the assignment,

3.2 Assist and communicate with the evaluators during the research phase.

Step 1 Initial considerations	**Step 2** Preparation of evaluation tasks	**Step 3** The evaluation research phase	**Step 4** Reporting and dissemination	**Step 5** Management response

3.1

The inception phase

In the tender document, the evaluators present their proposal for carrying out the evaluation. However, when preparing the tender document, evaluators are usually not in a position to formulate a fully operational evaluation plan. It may therefore be necessary to include an inception period in the research phase.

Provision for an inception study should be made in the ToR. In many cases, the ToR explicitly require that the inception report must be accepted by the client before the evaluation can proceed to the next phase.

It should also be noted that an inception report can be a very small part of the assignment, comprising a few days work resulting in a short report or perhaps simply a fine-tuning meeting with the evaluation manager. In some cases, not even a small inception study like this may be required.

The scope and focus of the inception study vary from case to case, but the following are standard topics:

- ■ Remaining evaluability issues,
- ■ Interpretation of evaluation questions,
- ■ Methodology for data collection and analysis, and
- ■ The evaluation work plan.

These issues are briefly reviewed in the following sections.

Remaining evaluability issues

As mentioned in Section 2.4, evaluability questions cannot always be conclusively dealt with before the evaluation starts. The assessment may have to continue into the inception phase and beyond. When this is the case, such questions must be discussed in the inception report.

Interpretations of the evaluation questions

When the evaluation questions in the ToR are formulated in general terms, the evaluators must translate them into questions that are more directly researchable. For example, if the ToR ask about the efficiency of the intervention, the evaluators may have to provide external benchmarks that clarify what could or should count as an efficient use of resources in the particular case at hand. Likewise, the evaluators may have to reinterpret unclear goals and objectives to make them empirically verifiable. The inception report provides an early opportunity for evaluation managers and evaluators to ensure that their interpretations of the ToR are mutually consistent.

Methodology and work plan

If required by the evaluation manager during contract negotiations, the inception report should provide further information about the proposed methodology and work plan, beyond that presented in the tender documents. Also, if the evaluation questions are changed or further developed during the inception period, the methodology and work plan are likely to need similar adjustment. If the evaluation includes case studies and the selection of the cases is made by the evaluators, the procedures of selection should be clarified in the inception report.

CHECKLIST

for supervising the inception period

☐ Ensure that the inception study deals with evaluability questions that could not be addressed before the evaluators were contracted.

☐ Ensure that the inception period develops an operational methodology and work plan on the basis of those evaluation questions that are considered answerable.

☐ Decide on necessary changes in relation to the terms of reference and the evaluators' technical proposal and regulate these changes in an addendum to the contract for the assignment.

3.2 Supporting the evaluators during the research phase

The evaluation manager and the evaluators should strive to develop an effective working relationship. Both parties should emerge from the planning stage with a clear understanding of how the evaluation is to be carried out, who is to do what, what is to be produced and when delivery is expected.

Unexpected developments may occur even after the inception phase. When this happens, the ToR and the other documents governing the evaluation process may be open to conflicting interpretations that can only be resolved through informal discussion. To be able to deal with upcoming problems both parties must maintain an open and flexible attitude.

During the research phase the role of the evaluation manager is largely that of a broker. Providing background documents and letters of introduction, arranging site visits, booking interviews are only a few activities where evaluators may require support from the evaluation manager. When the results of the inception study are discussed, the evaluation manager is advised to enquire which assistance is needed.

CHECKLIST

for communication and practical assistance during the research phase

☐ Agree with the evaluators before or during the field visits which upcoming and strategic issues may require consultation.

☐ Ensure that all strategic decisions about the evaluation are taken by the evaluation manager, also during the research phase, through recurrent communication with the evaluators.

☐ Ensure that the evaluators receive adequate help with practical matters during the field visits, for example documents, booking interviews and preparing site visits.

Reporting and Dissemination

In this fourth step of the evaluation process, the most important tasks of the evaluation manager are to:

4.1 Examine the evaluators' draft report against the contract, including any addenda, and formulate a written response. Make sure that other key stakeholders are also invited to respond to the draft report,

4.2 Make sure that the final report accommodates stakeholder comments on the draft report,

4.3 Disseminate the results of the evaluation to all interested parties, in particular the intended users, and

4.4 Facilitate publication of the final report in Sida's evaluation series, as well as on Sida's website.

Step 1	Step 2	Step 3	Step 4	Step 5
Initial considerations	Preparation of evaluation tasks	The evaluation research phase	Reporting and dissemination	Management response

4.1

Checking the draft report

This task consists of two main parts:

- Make sure that the report fulfils agreed presentation format and content as well as language and style.
- Make sure that the report contains a qualitatively satisfactory response to the evaluation questions in the ToR.

Virtually all evaluations are presented as written reports. The main objective of the report is to convey the results of the evaluation in a way that corresponds to the information needs of the intended users of the evaluation.

Evaluation users may have little patience with difficult and time-consuming language. An evaluation report that stimulates the readers' interest, matches their decision-making and learning requirements, and economises with their time, enhances the overall value of the evaluation.

The following is standard advice for effective reporting:

- Present main findings and conclusions up-front and use the rest of the report for more detailed analyses and presentation of findings,

- Focus on readers' expectations with regard to the object of evaluation. When learning is the purpose, highlight the unexpected and the problematic and do not dwell on matters of limited value to intended readers,

- Make sure that the overall structure of the report is clear and easy to understand. Describe the intervention logic, explain the evaluation questions, and be explicit about evaluation criteria and standards of performance,

- Present negative findings constructively. Be frank about shortcomings and mistakes, but avoid blame,

- As far as possible, avoid jargon and difficult technical terms,

- In evaluations commissioned by Sida we should try to follow the terminological conventions presented in this handbook and in the OECD/DAC Evaluation Glossary (Annex C),

- Use a consistent and conventional system for footnotes and references in the text,

- Explain abbreviations and consider listing them in a separate glossary,

- Use tables and figures to facilitate understanding.

Sida recommends a model format for evaluation reports (Annex B). Regulating the structure and content of seven standard chapters, this format is intended both to facilitate writing reports by evaluators and checking reports by evaluation managers and others. The format is not compulsory, but it should be used unless there is good reason for doing otherwise. The evaluators should consult with Sida's evaluation manager before adopting a different framework.

Report structure

EXECUTIVE SUMMARY
Summary of the evaluation, with particular emphasis on main findings, conclusions, lessons learned and recommendations.

INTRODUCTION
Presentation of the evaluation's purpose, questions and main findings.

THE EVALUATED INTERVENTION
Description of the evaluated intervention, and its purpose, logic, history, organisation and stakeholders.

FINDINGS
Factual evidence relevant to the questions asked by the evaluation and interpretations of such evidence.

EVALUATIVE CONCLUSIONS
Assessments of intervention results and performance against given evaluation criteria and standards of performance.

LESSONS LEARNED
General conclusions with a potential for wider application and use.

RECOMMENDATIONS
Actionable proposals regarding improvements of policy or management addressed to the client of the evaluation or other intended users.

ANNEXES
Terms of reference, methodology for data collection and analysis, references, etc.

Making sure that the evaluation report satisfies basic formal requirements is relatively simple. First, we examine if the report is organised as agreed and that no part is missing. Next, we make sure that all questions raised in the ToR have been covered and that the text is clear and succinct. While more thorough than the preliminary perusal of the report, this second task is not especially complicated. The points above can serve as a checklist.

Assessing the quality of response to evaluation questions can be more difficult. Exactly what constitutes acceptable quality in this regard varies from case to case.

The following are general quality criteria:

CLEAR STATEMENT OF THE EVALUATION QUESTIONS
The report should contain a clear restatement of the questions raised in the ToR so that readers will understand how the information in the report should be interpreted. Revisions of the original questions made in the course of the study should be duly noted.

CLEAR PRESENTATION OF CRITERIA AND STANDARDS OF PERFORMANCE
The report should clearly present evaluation criteria and standards of performance. The grounds for value judgements made in the report should be explicitly stated.

TRANSPARENT ACCOUNT OF RESEARCH METHODS

The report should include an account of sources of data and methods of data collection to help readers assess the likely accuracy of facts and figures.

JUSTIFIED CONCLUSIONS

It should be possible for readers to follow each step of the argument leading from question to answer. Supporting evidence should be clearly presented and alternative explanations of findings explicitly considered and eliminated. To help readers assess the quality of arguments, the report should contain a description of the logic of the intervention.

IMPARTIAL REPORTING

The perspectives of all major stakeholder groups should be impartially reflected in the report. The report should not give precedence to any particular perspective or point of view without saying so. It must cover both strengths and weaknesses, and should not be written in a manner that suggests that it is totally unbiased and represents the final truth.

CLEAR STATEMENT OF LIMITATIONS

All studies have limitations. They are limited with regard to scope and coverage as well as regarding depth of analysis. Unless these limitations are noted in the text, however, they may be difficult for readers to detect. Therefore, an account of major limitations should normally be included in reports.

Here a warning against encroaching on the evaluators' domain is warranted. As evaluation managers we try to make sure that the evaluation report conforms to the ToR and generally accepted quality standards for evaluations. While doing this, however, we must respect the evaluators' independence.

□ Check that the report meets the formal requirements stipulated by the contract and any contractual addenda.

□ Ensure that the report conforms to the agreed format for the structure and main contents of evaluation reports. In evaluations initiated by Sida, this is usually the format recommended by Sida (see Annex B).

□ Check that the report is well written and that it provides a qualitatively satisfactory response to the evaluation questions.

CHECKLIST

for the examination of the draft report

4.2

Accommodating stakeholder comments

The evaluation manager should make sure that the draft report is shared with the intended users of the evaluation and representatives of other key stakeholder groups. The consulted persons should have sufficient time to read and comment. Even if the impartiality of the evaluators is not disputed, comments from stakeholders can often help the evaluators correct factual errors and add important information. Stakeholders are normally in a far better position to identify gaps and mistakes than the evaluation manager.

It is important that the participatory perspective is not lost when the research phase is over. Persons who contribute to an evaluation in which they have a stake naturally want assurances that they have not been misinterpreted. For practical reasons it is often impossible to consult each and every person who has participated in a study. Yet, representatives of the intended users of the evaluation and other key stakeholders should be given an opportunity to comment before the study is finalised. Asking stakeholders to assess the draft report is not just a moral requirement. It is also ensures that the evaluation will be as useful as possible.

Sida's Intranet, e-mail and electronic project rooms may be used to facilitate discussions about the report. It is also a good idea to arrange early face-to-face meetings between the evaluation manager, major stakeholder groups and the evaluators. In such meetings, alternative views about the evaluation can be debated and put to test. In many cases, meetings held to discuss draft reports are more productive than meetings where the final results are presented.

When the revised draft report has been submitted, the evaluation manager should check that relevant stakeholder comments have been duly taken into account.

CHECKLIST

for accommodating stakeholder comments in the final report

☐ Solicit comments from intended users and other key stakeholders.

☐ When possible, consider arranging a meeting between major stakeholders and the evaluators to discuss the draft report.

☐ Check if stakeholder comments are adequately reflected in the report.

Dissemination of evaluation results

The dissemination of evaluation results must always, as with any successful communication strategy, be tailored to the audiences' needs and interests. It is important not to lose sight of the fact that a sound dissemination strategy tends to enhance the overall usefulness of the evaluation.

Some users may require information specifically targeted for decision-making with regard to the evaluated intervention. Others may seek to apply the lessons learned from the evaluation to other interventions and circumstances. Still others merely have an overall policy interest in the study.

The evaluation manager should work out a strategy for communicating the results of the evaluation in consultation with partners and other stakeholders. The discussions about communication and dissemination should not be postponed to the very end, when the study has been completed. Rather, they should be an integral part of the planning of the evaluation.

Apart from report writing, there is a wide range of options available to disseminate evaluation results. Meetings, seminars, workshops, conferences, media presentations are just some of the options. Information technology and software for special presentations may also be considered.

☐ Discuss with relevant stakeholders to whom, when and how the results of the evaluation should be disseminated, and implement the dissemination plan accordingly.

☐ Consider a range of dissemination options, including dialogue with partners, meetings, seminars, workshops, and any other kind of relevant and effective communications strategy.

☐ Make sure that the dissemination is tailored to the specific needs, interests and information requirements of individual audiences.

☐ Focus dissemination efforts on the intended users of the evaluation, and other groups that can be expected to make effective use of the evaluation.

CHECKLIST

for dissemination of evaluation results

Publishing the evaluation report

When the final report has been approved, the evaluation manager should make sure that it can be published by Sida's Department for Evaluation and Internal Audit (UTV) in one of the two report series Sida Evaluations and Sida Studies in Evaluation, as well as on Sida's website on the Internet. It is Sida policy that evaluations should be made accessible to the international development community as a whole.

The evaluation manager should also make sure that all statistical details of Sida Evaluations Data Worksheet are delivered to UTV together with the full report.

The final responsibilities of the evaluation manager are to decide if the report should be translated into another language, to produce a mailing list of people and institutions that should receive the report, and to ensure that copies are dispatched in print or electronically without undue delay.

CHECKLIST

**for publication of
the evaluation report**

☐ Check the guidelines for the publication and dissemination of evaluation reports and the instructions regarding evaluation statistics in Sida Templates/UTV on Sida's intranet.

☐ Check that the final report received from the consultant is ready for publication and that it has been delivered along with the required statistical details about the assignment.

☐ Consider translation of the report into relevant languages.

☐ Make a list of people and institutions that should receive the report, and dispatch printed or electronic copies as soon as possible.

Management Response

This fifth and last step of the evaluation process consists of two tasks:

5.1 Make sure that the evaluation is submitted for management response, and assist the concerned embassy or department in interpreting the evaluation report,

5.2 Facilitate the use of the evaluation as an input to the dialogue between Sida, its co-operation partners and other stakeholders.

Step 1	Step 2	Step 3	Step 4	Step 5
Initial considerations	Preparation of evaluation tasks	The evaluation research phase	Reporting and dissemination	Management response

Both these tasks may be important for the overall success of the evaluation process. An evaluation is of little value if its conclusions, lessons and recommendations are not properly understood by its intended users and other interested parties.

5.1

Assisting in the formulation of Sida's management response

To ensure that evaluations are taken into account by concerned departments and embassies, and that their response is well documented, Sida has developed a system for management response. As stated in Sida's Evaluation Policy, all Sida-financed evaluations should be followed up by a management response outlining the concerned embassy's or department's conclusions with regard to the evaluation. In most cases, the management response includes an action plan and time schedule, indicating what will be done as a result of the study and when it will be done. If the embassy or department rejects some of the recommendations of the evaluation, or plans to implement certain recommendations in different ways than those proposed by the evaluators, the reasons for such modification should be clearly explained.

As a rule, a Sida management response consists of three main parts:

- An overall assessment from Sida's point of view of the relevance, accuracy and usefulness of the evaluation and its findings.

- A point-by-point response to the recommendations and/ or main findings. Are the recommendation accepted or rejected? Will they prompt action? Do the findings and conclusions of the evaluation have any practical implications beyond those raised in the recommendations?

- A summary action plan with a completion date for each action. If it is decided that no measures should be taken in response to the evaluation, an action plan is obviously not necessary.

The evaluation manager is normally actively involved in the formulation of the management response. When the evaluation manager is identical with the programme officer in charge of Sida's support to the evaluated intervention, which is often the case, he or she would usually draft the response. In other cases, the evaluation manager may usefully assist the responsible embassy or department in interpreting the evaluation and its findings.

CHECKLIST

for assisting the formulation of a management response

- ☐ In consultation with the concerned embassy or department, draft a management response to the evaluation. Alternatively, assist those who are drafting the management response to interpret the evaluation report.

- ☐ If necessary, make sure that the management response is complete and satisfies Sida's requirements.

5.2

Facilitating dialogue with co-operation partners

This manual cannot provide instructions for how Sida's co-operation partners and other stakeholders should respond to evaluation results. If the evaluation report contains recommendations addressed to Sida's co-operation partner, they will be processed in accordance with the partner's own procedures.

As a specialist on the evaluation in question, however, the evaluation manager is often well equipped to identify issues that ought to be raised in discussions between Sida and the co-operation partner. As noticed elsewhere, Sida frequently uses evaluation results as inputs to the dialogue with its partners.

CHECKLIST

for facilitating the dialogue with co-operation partners

☐ Remember to make relevant dialogue use of the evaluation also after the evaluation process has been concluded and the management response is implemented.

Annex A

Format for Terms of Reference

The format below is intended to help guide the structure and main content of the terms of reference (ToR) for Sida financed evaluations:

EVALUATION PURPOSE

Details about the intended use of the evaluation, including any operative decisions the evaluation is expected to feed into.

INTERVENTION BACKGROUND

Details about the intervention logic, as well as the structure and substance of the activities and outputs carried out and delivered.

STAKEHOLDER INVOLVEMENT

Account of how stakeholders are expected to participate in the research, reporting and dissemination activities of the evaluation.

EVALUATION QUESTIONS

List of evaluation questions, based on a set of criteria, policy issues and performance standards that suit the evaluation purpose.

RECOMMENDATIONS AND LESSONS

Clarification of what kinds of recommendations and lessons the evaluation is intended to provide.

METHODOLOGY

Directives regarding the methodology to be used by the evaluators for data collection and analysis.

WORK PLAN AND SCHEDULE

Indications of e.g. what sites should be visited and how evaluators' time should be divided between field and reporting phases.

REPORTING

Specification of what reports should be delivered and when, and the evaluators' role in follow up seminars and workshops.

EVALUATION TEAM

The qualifications of the evaluators, such as evaluation skills, sector knowledge and socio-cultural competence.

By following this model format, the evaluation ToR are likely to cover many aspects that are necessary for successfully guiding the evaluation process. Still, note that the evaluation manager in many cases cannot be expected to gather all information required by this format before the ToR are finalised.

Evaluation purpose

In this first section of the ToR, the evaluation purpose is defined. Remember to specify a purpose that reflects the intended use of the evaluation. If this intended use concerns operative decisions for project cycle management and dialogue, make sure to elaborate also on the particular context of such decision-making. State who the intended users of the evaluation are. The most typical users of Sida-financed evaluations are Sida departments, Swedish embassies, and co-operation partners. See Section 2.2 of the second part of the manual for more details about the evaluation purpose.

Intervention background

In this section, the intervention selected for evaluation is described in order to give evaluators a better understanding of what they should evaluate. In particular, the logic of the intervention should be described, including: the goal structure from inputs and activities up to the overall goal of poverty reduction; the indicators that provide specificity to the intervention logic; and the internal and external assumptions on which the logic is based.

It is essential that the ToR are explicit about the time periods and intervention phases that the evaluation should focus on. Also, if the intervention is complex, and if the evaluation manager and concerned stakeholders have agreed on a particular focus on certain intervention components, the details about this focus need to be spelled out in the ToR. Note that the description of the intervention logic normally needs to deal only with those components on which the evaluation focuses.

If the evaluation manager does not have access to full information on the selected intervention when preparing the evaluation, which may often be the case, the ToR should indicate how and when additional gathering of preparatory information should be undertaken as part of the evaluation assignment. See Section 3.1 of the manual for more details in this respect.

Stakeholder involvement

This section summarises the stakeholder analysis made in the preparatory step of the evaluation process. It also identifies all agreements on how different stakeholder groups are to participate in and contribute to the research, reporting and dissemination phases of the evaluation.

In many cases, partners and other stakeholders will also participate during the implementation of the study. The ToR should specify how the evaluators are expected to interact with different groups of stakeholders during the process.

Evaluation questions

This section lists the evaluation questions that should be answered by the evaluation. The list should be based on a feasible choice of evaluation criteria and policy issues that suits the evaluation purpose. Since evaluation questions tend to be difficult and time-consuming to answer with adequate precision and reliability, it is often a good idea to limit the number of questions in order to allow them to be addressed in depth.

Evaluation questions can often not be decided before it is determined whether they really can be answered as intended. If this evaluability assessment has not been carried out when the ToR are being finalised, the ToR need to indicate if and how this assessment will be undertaken and how, later in the process, it will be decided what questions the evaluation should address.

Recommendations and lessons

While it is impossible during the planning stage of the evaluation process to foresee answers to specific evaluation questions, the ToR should indicate what kinds of recommendations and lessons that the evaluators should provide. For example, are we interested in recommendations and lessons about a particular form of co-operation, a particular socio-economic context, or a certain policy? Even if the kinds of recommendations and lessons that are required given a specific evaluation purpose are self-evident, the evaluation manager is generally advised to be explicit in this respect.

Methodology

Normally, the evaluators are responsible for the research methods. The chosen methods should be described and justified in relation to possible alternatives in the tender documents, or in the inception report produced at an early stage of the assignment. Instructions for when and how a discussion about methodology is needed should, however, always be given in the ToR.

Work plan and schedule

In this section, the ToR give instructions for the practical aspects of the evaluation process, such as what parts of the evaluation should be carried out in Sweden and in partner countries, and what partner countries and sites that should be visited by the evaluators.

As in the case of evaluation methodology, details about the evaluation work plan is often something that needs to be elaborated in the tender and the inception phase of the assignment. However, it may be necessary for the ToR to indicate when the assignment is to be concluded, and, roughly, how the evaluators'

time should be divided between the inception, field and reporting phases of the evaluation.

Reporting

This final section specifies the reports that should be delivered under the evaluation contract, such as inception reports, and draft and final reports. The consultant should be instructed to adhere to the terminological conventions of the OECD/DAC Glossary on Evaluation and Results-Based Management as far as possible. The section should further specify delivery dates for the reports, and the evaluators' roles in follow up activities such as seminars and workshops.

That the evaluation report should consider the report format presented in Annex B of this manual, and that a completed Sida Evaluations Data Work Sheet should be presented along with the report also need to be clarified. It should be explicitly noted that evaluation reports will be assessed against standard quality criteria for evaluation reporting, such as those described in this manual.

Evaluation team

This section defines the necessary qualifications of the evaluation team and individual team members, for example in terms of evaluation skills, country and sector experience, and social and cultural competence. When the evaluation is intended to serve a purpose of accountability, a requirement that the evaluators should be independent of the evaluated activities and have no stake in the outcome of the evaluation must be inserted in the ToR.

For some further notes on the writing of terms of reference, consult the Evaluation Manual, section 2.6.

Annex B

Format for Sida Evaluation Reports

This format is intended to help guide the structure and main contents of evaluation reports commissioned by Sida. It is not compulsory, but should be used if there is no particular reason for doing otherwise.

Report structure

EXECUTIVE SUMMARY
Summary of the evaluation, with particular emphasis on main findings, conclusions, lessons learned and recommendations.

INTRODUCTION
Presentation of the evaluation's purpose, questions and main findings.

THE EVALUATED INTERVENTION
Description of the evaluated intervention, and its purpose, logic, history, organisation and stakeholders.

FINDINGS
Factual evidence, data and observations that are relevant to the specific questions asked by the evaluation.

EVALUATIVE CONCLUSIONS
Assessment of the intervention and its results against given evaluation criteria, standards of performance and policy issues.

LESSONS LEARNED
General conclusions that are likely to have a potential for wider application and use.

RECOMMENDATIONS
Actionable proposals to the evaluation's users for improved intervention cycle management and policy.

ANNEXES
Terms of reference, methodology for data gathering and analysis, references, etc.

By following a uniform format, evaluation reports tend to be easier to read and use. The format also facilitates syntheses of different reports for broader learning purposes, such as in Sida's results analyses for the development of new country strategies.

The format may be included as an annex to the contract with the consultant, thus providing early instructions how the report may be prepared. However, note that Sida's Evaluation

Manual contains further guidance about reporting, and that the evaluator is well advised to take a look at the manual as a whole. The present format is found in Templates/UTV on Sida's intranet.

Executive summary

The executive summary provides a synopsis of the evaluation and its purpose, emphasising main findings, evaluative conclusions, recommendations and lessons learned. Descriptions of methodology should be kept to a minimum.

The summary should be self-contained and self-explanatory. Special care should be taken to prepare the executive summary, as it is may be the only part of the report that some people have time to read.

Introduction

The introduction presents the background and overall purpose of the evaluation, including how and by whom it is intended to be used, as well as the evaluation criteria employed and the key questions addressed. It also outlines the structure of the report and provides guidance to readers.

The evaluated intervention

This chapter describes the main characteristics of the evaluated intervention and its location, history, organisation and stakeholders. It should cover the focal problem addressed by the evaluated intervention, the objectives of the invention and its logic of cause and effect. A description of activities carried out and key outputs delivered should be included.

The chapter should also cover the policy and development context of the evaluated intervention, including the assumptions about external factors that were part of intervention planning. When preparing the chapter, the evaluators should summarize the findings and conclusions of any earlier evaluations of the same intervention.

Findings

Findings are empirical data and inferences from such data that the evaluators present as evidence relevant to the evaluation questions. They are the facts of the matter, in other words.

In the findings chapter, this body of evidence is systematically presented so that readers can form their own opinion about the strengths and weakness of the conclusions of the evaluation. The quality of the findings – their accuracy and relevance – should be assessed with reference to standard criteria of reliability and validity.

Evaluative conclusions

Evaluative conclusions are the evaluators' concluding assessments of the intervention against given evaluation criteria, performance standards and policy issues. They provide answers as to whether the intervention is considered good or bad, and whether the results are found positive or negative.

Note that the distinction between findings and evaluative conclusions is somewhat artificial. Evaluative conclusions are often best presented together with the underlying findings on which they are based. In many cases, it makes sense to combine the presentation of findings and evaluative conclusions in one chapter.

Lessons learned

Lessons learned are findings and conclusions that can be generalised beyond the evaluated intervention.

In formulating lessons, the evaluators are expected to examine the intervention in a wider perspective and put it in relation to current ideas about good and bad practice.

Recommendations

Recommendations indicate what actions the evaluators believe should be taken on the basis of the evaluation. Recommendations to Sida may cover the whole spectrum of aid management, including resource allocation, financing, planning, implementation, and monitoring and evaluation.

Recommendations should always identify their respective addressees and be tailored to the specific needs and interests of each addressee. They should be simply stated and geared to facilitate implementation.

Annex on methodology

The report should include an annex describing how the evaluation was carried out. The annex should cover standard methodology topics, including research design, sampling and data collection methods and analytical procedures. It should discuss the limitations of the selected methods as well as their strengths.

Annex C

Glossary of Key Terms in Evaluation and Results-Based Management

This glossary has been developed by the OECD/DAC Network on Development Evaluation (formerly The DAC Working Party on Aid Evaluation). Completed in 2002, it is available in several languages, including French and Spanish. All the different versions can be downloaded from the OECD website (www.oecd.org).

The manual follows the terminology of the glossary on most points. The differences, less than a dozen in all, are noted below, entry by entry. All the inserted notes have been marked with an asterisk (*).

> The Development Assistance Committee (DAC) Network on Development Evaluation is an international forum where representatives of the evaluation departments of bilateral and multilateral development organisations meet to share experience and to improve evaluation practice and strengthen its use as an instrument for development co-operation policy. It operates under the aegis of DAC and presently consists of 30 representatives from OECD member countries and multilateral development agencies (Australia, Austria, Belgium, Canada, Denmark, European Commission, Finland, France, Germany, Greece, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States, World Bank, Asian Development Bank, African Development Bank, Inter-American Development Bank, European Bank for Reconstruction).

A **ACCOUNTABILITY** Obligation to demonstrate that work has been conducted in compliance with agreed rules and standards or to report fairly and accurately on performance results vis-à-vis mandated roles and/or plans. This may require a careful, even legally defensible, demonstration that the work is consistent with the contract term.

Note: Accountability in development may refer to the obligations of partners to act according to clearly defined responsibilities, roles and performance expectations, often with respect to the prudent use of resources. For evaluators, it connotes the responsibility to provide accurate, fair and credible monitoring reports and performance assessments. For public sector managers and policy-makers, accountability is to taxpayers/citizens.

A   **ACTIVITY** Actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilised to produce specific outputs.

Related term: Development intervention.

ANALYTICAL TOOLS Methods used to process and interpret information during an evaluation.

APPRAISAL An overall assessment of the relevance, feasibility and potential sustainability of a development intervention prior to a decision of funding.

Note: In development agencies, banks, etc., the purpose of appraisal is to enable decision-makers to decide whether the activity represents an appropriate use of corporate resources.

Related term: Ex-ante evaluation.

ASSUMPTIONS Hypotheses about factors or risks which could affect the progress or success of a development intervention.

Note: Assumptions can also be understood as hypothesized conditions that bear on the validity of the evaluation itself, e.g. about the characteristics of the population when designing a sampling procedure for a survey. Assumptions are made explicit in theory based evaluations where evaluation tracks systematically the anticipated results chain.

ATTRIBUTION The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention.

Note: Attribution refers to that which is to be credited for the observed changes or results achieved. It represents the extent to which observed development effects can be attributed to a specific intervention or to the performance of one or more partner taking account of other interventions, (anticipated or unanticipated) confounding factors, or external shocks.

AUDIT An independent, objective assurance activity designed to add value and improve an organisation's operations. It helps an organisation accomplish its objectives by bringing a system-atic, disciplined approach to assess and improve the effectiveness of risk management, control and governance processes.

Note: A distinction is made between regularity (financial) auditing, which focuses on compliance with applicable statutes and regulations; and performance auditing, which is concerned with relevance, economy, efficiency and effectiveness. Internal auditing provides an assessment of internal controls undertaken by a unit reporting to management while external auditing is conducted by an independent organisation.

B **BASE-LINE STUDY** An analysis describing the situation prior to a development intervention, against which progress can be assessed or comparisons made.

BENCHMARK Reference point or standard against which performance or achievements can be assessed.
   Note: A benchmark refers to the performance that has been achieved in the recent past by other comparable organisations, or what can be reasonably inferred to have been achieved in the circumstances.

BENEFICIARIES The individuals, groups, or organisations, whether targeted or not, that benefit, directly or indirectly, from the development intervention.
   Related terms: Reach, target group.

C **CLUSTER EVALUATION** An evaluation of a set of related activities, projects and/or programs.

CONCLUSIONS Conclusions point out the factors of success and failure of the evaluated intervention, with special attention paid to the intended and unintended results and impacts, and more generally to any other strength or weakness. A conclusion draws on data collection and analyses undertaken, through a transparent chain of arguments.

COUNTERFACTUAL The situation or condition which hypo-thetically may prevail for individuals, organisations, or groups were there no development intervention.

COUNTRY PROGRAM EVALUATION/COUNTRY ASSISTANCE EVALUATION Evaluation of one or more donor's or agency's portfolio of development interventions, and the assistance strategy behind them, in a partner country.

D **DATA COLLECTION TOOLS** Methodologies used to identify information sources and collect information during an evaluation.
   Note: Examples are informal and formal surveys, direct and participatory observation, community interviews, focus groups, expert opinion, case studies, literature search.

DEVELOPMENT INTERVENTION An instrument for partner (donor and non-donor) support aimed to promote development.
   Note: Examples are policy advice, projects, programs.

D **DEVELOPMENT OBJECTIVE** Intended impact contributing to physical, financial, institutional, social, environmental, or other benefits to a society, community, or group of people via one or more development interventions.

E **ECONOMY** Absence of waste for a given output.
Note: An activity is economical when the costs of the scarce resources used approximate the minimum needed to achieve planned objectives.

EFFECT Intended or unintended change due directly or indirectly to an intervention.
Related terms: Results, outcome.

EFFECTIVENESS The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.
Note: Also used as an aggregate measure of (or judgment about) the merit or worth of an activity, i.e. the extent to which an intervention has attained, or is expected to attain, its major relevant objectives efficiently in a sustainable fashion and with a positive institutional development impact.
Related term: Efficacy.

EFFICIENCY A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.

EVALUABILITY Extent to which an activity or a program can be evaluated in a reliable and credible fashion.
Note: Evaluability assessment calls for the early review of a proposed activity in order to ascertain whether its objectives are adequately defined and its results verifiable.

* Ideally, an evaluability assessment should be made when a development intervention is planned. However, evaluability must also be assessed again as a prelude to evaluation.

E **EVALUATION** 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.

Note: Evaluation in some instances involves the definition of appropriate standards, the examination of performance against those standards, an assessment of actual and expected results and the identification of relevant lessons.

Related term: Review.

* The definition of evaluation in Sida's Evaluation Policy differs only slightly from the DAC definition: An evaluation is a careful and systematic retrospective assessment of the design, implementation, and results of development activities.

EX-ANTE EVALUATION An evaluation that is performed before implementation of a development intervention.

Related terms: Appraisal, quality at entry.

EX-POST EVALUATION Evaluation of a development intervention after it has been completed.

Note: It may be undertaken directly after or long after completion. The intention is to identify the factors of success or failure, to assess the sustainability of results and impacts, and to draw conclusions that may inform other interventions.

EXTERNAL EVALUATION The evaluation of a development intervention conducted by entities and/or individuals outside the donor and implementing organisations.

F **FEEDBACK** The transmission of findings generated through the evaluation process to parties for whom it is relevant and useful so as to facilitate learning. This may involve the collection and dissemination of findings, conclusions, recommendations and lessons from experience.

FINDING A finding uses evidence from one or more evaluations to allow for a factual statement.

F **FORMATIVE EVALUATION** Evaluation intended to improve performance, most often conducted during the implementation phase of projects or programs.

Note: Formative evaluations may also be conducted for other reasons such as compliance, legal requirements or as part of a larger evaluation initiative.

Related term: Process evaluation.

G **GOAL** The higher-order objective to which a development intervention is intended to contribute.

Related term: Development objective.

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

* As noted in Chapter 2, the word is widely used in a more comprehensive sense that includes both short and long-term effects. In this manual, it is used in the broader as well as in the more narrow sense defined by the Glossary.

INDEPENDENT EVALUATION An evaluation carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention.

Note: The credibility of an evaluation depends in part on how independently it has been carried out. Independence implies freedom from political influence and organisational pressure. It is characterised by full access to information and by full autonomy in carrying out investigations and reporting findings.

* This manual distinguishes between two types of independent evaluation. In the one case the evaluators are independent of the evaluated activities and have no stake in the outcome of the study. In the other case, there is a further requirement that the evaluation is also commissioned by an organisation that is independent of the evaluated activities.

INDICATOR Quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor.

INPUTS The financial, human, and material resources used for the development intervention.

I   **INSTITUTIONAL DEVELOPMENT IMPACT** The extent to which an intervention improves or weakens the ability of a country or region to make more efficient, equitable, and sustainable use of its human, financial, and natural resources, for example through: (a) better definition, stability, transparency, enforceability and predictability of institutional arrangements and/or (b) better alignment of the mission and capacity of an organisation with its mandate, which derives from these institutional arrangements. Such impacts can include intended and unintended effects of an action.

   INTERNAL EVALUATION Evaluation of a development intervention conducted by a unit and/or individuals reporting to the management of the donor, partner, or implementing organisation.
     Related term: Self-evaluation.

J   **JOINT EVALUATION** An evaluation to which different donor agencies and/or partners participate.
     Note: There are various degrees of "jointness" depending on the extent to which individual partners co-operate in the evaluation process, merge their evaluation resources and combine their evaluation reporting. Joint evaluations can help overcome attribution problems in assessing the effectiveness of programs and strategies, the complementarity of efforts supported by different partners, the quality of aid co-ordination, etc.

L   **LESSONS LEARNED** Generalisations based on evaluation experiences with projects, programs, or policies that abstract from the specific circumstances to broader situations. Frequently, lessons highlight strengths or weaknesses in preparation, design, and implementation that affect performance, outcome, and impact.

* As the term is understood in this manual, the degree of generalisation of a lesson varies from case to case. As the conditions for development co-operation vary, illuminating attempts at generalisation are often restricted to a particular type of context or mode of intervention.

L **LOGICAL FRAMEWORK (LOGFRAME)** Management tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators, and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention.

Related term: Results-based management.

* It should be noted that logframe analysis (LFA) is one of several closely related types of analyses that focus on the chain of cause and effect underlying the evaluated intervention. Programme logic models, theories of action, performance frameworks, project theories, and development hypotheses are all members of the same family as the logframe. In this manual, the term intervention logic serves as a blanket term.

M **META-EVALUATION** The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators.

MID-TERM EVALUATION Evaluation performed towards the middle of the period of implementation of the intervention.

Related term: Formative evaluation.

MONITORING A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds.

Related term: Performance monitoring, indicator.

O **OUTCOME** The likely or achieved short-term and medium-term effects of an intervention's outputs.

Related terms: Result, output, impact, effect.

* Among evaluators the word outcome is also frequently used in a general sense where it is more or less synonymous with the word effect. When it is used in this sense, distinctions are made between short, medium, and long-term outcomes.

OUTPUTS The products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes.

P **PARTICIPATORY EVALUATION** Evaluation methods through which representatives of agencies and stakeholders (including beneficiaries) work together in designing, carrying out and interpreting an evaluation.

* In this manual we distinguish between participatory evaluations and participatory evaluation methods. An evaluation may use participatory methods, and still not qualify as a fully participatory evaluation. This distinction is further clarified in Chapter 1.

PARTNERS The individuals and/or organisations that collaborate to achieve mutually agreed upon objectives.
   Note: The concept of partnership connotes shared goals, common responsibility for outcomes, distinct accountabilities and reciprocal obligations. Partners may include governments, civil society, non-governmental organisations, universities, professional and business associations, multilateral organisations, private companies, etc.

PERFORMANCE The degree to which a development intervention or a development partner operates according to specific criteria/standards/guidelines or achieves results in accordance with stated goals or plans.

PERFORMANCE INDICATOR A variable that allows the verification of changes in the development intervention or shows results relative to what was planned.
   Related terms: Performance monitoring, performance measurement.

PERFORMANCE MEASUREMENT A system for assessing performance of development interventions against stated goals.
   Related terms: Performance monitoring, indicator.

PERFORMANCE MONITORING A continuous process of collecting and analysing data to compare how well a project, program, or policy is being implemented against expected results.

PROCESS EVALUATION An evaluation of the internal dynamics of implementing organisations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these.
   Related term: Formative evaluation.

* As the term is understood in this manual, a process evaluation may also deal with outputs and other intermediary results.

P **PROGRAM EVALUATION** Evaluation of a set of interventions, marshaled to attain specific global, regional, country, or sector development objectives.

Note: A development program is a time bound intervention involving multiple activities that may cut across sectors, themes and/or geographic areas.

Related term: Country program/strategy evaluation.

PROJECT EVALUATION Evaluation of an individual development intervention designed to achieve specific objectives within specified resources and implementation schedules, often within the framework of a broader program.

Note: Cost benefit analysis is a major instrument of project evaluation for projects with measurable benefits. When benefits cannot be quantified, cost effectiveness is a suitable approach.

* As the concept is understood in this manual, there are many approaches to project evaluation. Cost-benefit analysis and analyses of cost-effectiveness are important tools for economic evaluation focussing on questions of efficiency.

PROJECT OR PROGRAM OBJECTIVE The intended physical, financial, institutional, social, environmental, or other development results to which a project or program is expected to contribute.

PURPOSE The publicly stated objectives of the development program or project.

Q **QUALITY ASSURANCE** Quality assurance encompasses any activity that is concerned with assessing and improving the merit or the worth of a development intervention or its compliance with given standards.

Note: Examples of quality assurance activities include appraisal, results-based management, reviews during implementation, evaluations, etc. Quality assurance may also refer to the assessment of the quality of a portfolio and its development effectiveness.

R **REACH** The beneficiaries and other stakeholders of a development intervention.

Related term: Beneficiaries.

RECOMMENDATIONS Proposals aimed at enhancing the effectiveness, quality, or efficiency of a development intervention; at redesigning the objectives; and/or at the reallocation of resources. Recommendations should be linked to conclusions.

R **RELEVANCE** The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies.
  Note: Retrospectively, the question of relevance often becomes a question as to whether the objectives of an intervention or its design are still appropriate given changed circumstances.

RELIABILITY Consistency or dependability of data and evaluation judgements, with reference to the quality of the instruments, procedures and analyses used to collect and interpret evaluation data.
  Note: Evaluation information is reliable when repeated observations using similar instruments under similar conditions produce similar results.

RESULT The output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention.
  Related terms: Outcome, effect, impact.

RESULTS CHAIN The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives beginning with inputs, moving through activities and outputs, and culminating in outcomes, impacts, and feedback. In some agencies, reach is part of the results chain.
  Related terms: Assumption, results framework.

RESULTS FRAMEWORK The program logic that explains how the development objective is to be achieved, including causal relationships and underlying assumptions.
  Related terms: Results chain, logical framework.

RESULTS-BASED MANAGEMENT (RBM) A management strategy focusing on performance and achievement of outputs, outcomes and impacts.
  Related term: Logical framework

REVIEW An assessment of the performance of an intervention, periodically or on an ad hoc basis.
  Note: Frequently "evaluation" is used for a more comprehensive and/or more in-depth assessment than "review". Reviews tend to emphasise operational aspects. Sometimes the terms "review" and "evaluation" are used as synonyms.
  Related term: Evaluation.

R **RISK ANALYSIS** An analysis or an assessment of factors (called assumptions in the logframe) affect or are likely to affect the successful achievement of an intervention's objectives. A detailed examination of the potential unwanted and negative consequences to human life, health, property, or the environment posed by development interventions; a systematic process to provide information regarding such undesirable consequences; the process of quantification of the probabilities and expected impacts for identified risks.

S **SECTOR PROGRAM EVALUATION** Evaluation of a cluster of development interventions within one country or across countries, all of which contribute to the achievement of a specific development goal.
   Note: A sector includes development activities commonly grouped together for the purpose of public action such as health, education, agriculture, transport etc.

SELF-EVALUATION An evaluation by those who are entrusted with the design and delivery of a development intervention.

STAKEHOLDERS Agencies, organisations, groups or individuals who have a direct or indirect interest in the development intervention or its evaluation.

SUMMATIVE EVALUATION A study conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth of the program.
   Related term: Impact evaluation.

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.

T **TARGET GROUP** The specific individuals or organisations for whose benefit the development intervention is undertaken.

TERMS OF REFERENCE Written document presenting the purpose and scope of the evaluation, the methods to be used, the standard against which performance is to be assessed or analyses are to be conducted, the resources and time allocated, and reporting requirements. Two other expressions sometimes used with the same meaning are "scope of work" and "evaluation mandate".

T   **THEMATIC EVALUATION** Evaluation of a selection of develop-
ment interventions, all of which address a specific development
priority that cuts across countries, regions, and sectors.

TRIANGULATION The use of three or more theories, sources
or types of information, or types of analysis to verify and
substantiate an assessment.
   Note: By combining multiple data-sources, methods, analyses
or theories, evaluators seek to overcome the bias that comes
from single informants, single-methods, single observer or single
theory studies.

V   **VALIDITY** The extent to which the data collection strategies and
instruments measure what they purport to measure.