Sida Studies in Evaluation 97/1

Using the Evaluation Tool

A survey of conventional wisdom and common practice at Sida

Jerker Carlsson
Kim Forss
Karin Metell
Lisa Segnestam
Tove Strömberg

Department for Evaluation
and Internal Audit

Using the Evaluation Tool

A survey of conventional wisdom and common practice at Sida

Jerker Carlsson
Kim Forss
Karin Metell
Lisa Segnestam
Tove Strömberg

Sida Studies in Evaluation 97/1

Department for Evaluation
and Internal Audit

Sida Studies in Evaluation is a series concerned with conceptial and methodological issues in the evaluation of development cooperation It is published by Sida's Department for Evaluation and Internal Audit. Reports may be ordered from:

Biståndsforum, Sida
S-105 25 Stockholm
Phone (+46) 8 698 57 22
Fax (+46) 8 698 56 38

Authors Jerker Carlsson, Kim Forss, Karin Metell, Lisa Segnestam, Tove Stromberg

The views and interpretations expressed in this report are the author's and do not necessarily reflect those of the Swedish International Development Cooperation Agency, Sida.

Sida Studies in Evaluation 97/1
Commissioned by Sida, Department for Evaluation and Internal Audit

Registration No UTV-1996-0025
Date of Final Report June 1997
Printed in Stockholm, Sweden 1997
ISSN 1402—215X

SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY
Address S-105 25 Stockholm, Sweden Office· Sveavagen 20, Stockholm
Telephone +46 (0)8-698 50 00. Telefax +46 (8)-20 88 64
Telegram sida stockholm. Telex 11450 sida sthlm Postgiro· 1 56 34–90
Homepage http //www sida se

Table of contents

1. Introduction

Purpose

One of the consequences of the reorganization of the Swedish aid organizations in 1995 was the creation of a new evaluation function. This function is now called the Department for Evaluation and Internal Audit (UTV). The new evaluation function differs from the old one in several respects: it has a more independent position as it reports to Sida's Board rather than to the Director-General, it has more resources, it has a wider mandate and it is merged with the internal audit function. Still, evaluations of development assistance are done by the sector and regional departments, although UTV has an overall responsibility for evaluation activities. Part of the tasks of UTV are thus to assist other parts of the organization in the professional conduct of evaluation work, to propagate for timely and well-conceived evaluation activities, and to review and monitor the overall quality of Sida´s evaluations.

When the five Swedish aid organization merged into one, they brought with them different practices and cultures with respect to evaluation. These differences were studied in a series of monographs by the former Secretariat for Analysis of Swedish Development Assistance (Stokke, 1994; Bruce and Finci, 1994, Carlsson and Jörnmark, 1994; Forss, 1994). However, these reports focus on the formalities of evaluation systems. They do not penetrate into a review of how the evaluation instrument as such is used, and they do not analyse the culture surrounding the use of the evaluation instrument. In addition, organizations usually change as their environment changes, and the cultures of the merging units would be expected to fuse into something entirely different as a result of confrontation with the practices of others. So, much remains to be known about the position, use and prospects of evaluations in the new organization.

One of the thematic areas for evaluation concerns organizational change and learning. The study which is presented here forms part of the activities in this field. Terms of reference for this particular study as well as for the thematic area are enclosed in annex 1. In short, the purpose of this study is to map the use of evaluations at Sida, by answering the following questions:

- How and why are evaluations initiated?
- How is the evaluation process managed, from the formulation of purpose, the decision to evaluate and the commissioning of a study?
- How are the results of this process used?

When we speak of evaluation, we use the definition which is given in Sida's evaluation policy (Sida, 1995):

> An evaluation is a careful and systematic ex-post assessment of the design, implementation and results of an activity in relation to its objectives Within the field of development assistance, subjects of evaluation may be one or several aspects of ongoing or completed projects, programmes, action plans, or policies An evaluation may also take the form of an assessment of one or several aspects of the capacity of Sida, or of organizations supported by Sida, to design and implement appropriate projects, programmes, action plans and policies

Method

The study builds on the application of two different methods, interviews, and application of a model for quality analysis of evaluation reports. The interviews were conducted with evaluation coordinators and with programme officers who have recently been in charge of an evaluation process. The quality analysis covers more than half of the evaluations completed during the past financial year (1995/96).

Interviews with evaluation coordinators

The first step in the analysis was to review the organization of the evaluation function, in particular by looking at the roles and responsibilities of the evaluation coordinators in each department of the organization. Sida was until 1 January 1997 organized in 14 departments.

As this study looks at the evaluations of development cooperation financed by Sida, it does not cover work undertaken by the departments for Administration, Information, Policy and Legal Services or Cooperation with Non-governmental Organization and Humanitarian Assistance. Nor does it cover the activities of UTV itself. This delimits the total population of objects to 9 departments. In each of these the evaluation coordinators were interviewed.

The interviews followed a structured format (the interview protocols are enclosed in annex 4). The format was designed by UTV, and the interviews were conducted by Karin Metell at UTV, who also analysed the responses and wrote the report which is enclosed in annex 2. (Each of the 9 interviews is printed and can be reviewed at UTV.)

Interviews with programme officers

The second step in the analysis was to conduct interviews with the programme officers at Sida who have recently been managing an evaluation process. The evaluation plan, which is compiled by UTV in cooperation with the departments, lists 117 evaluations that were to be conducted in the budget year 1995/96 (which covers 18 months). This plan was not fully realized, for a variety of reasons. At the end of 1996, only 54 of the originally planned 117 evaluations had been completed, along with a further 27 evaluations not on the list. The rest were either delayed, cancelled or transformed into something other than an evaluation. Of these 81 evaluations, 51 were published. And of these, only 30 were the subject of interviews for this study, conducted in November–December 1996. The short-fall of a large part of the evaluations is due to a variety of factors: some were not known to UTV at the time (the full list was not available until a follow-up was completed in January 1997); in several cases one programme officer was in charge of several evaluations, whereas the interview format only foresaw one evaluation per interview; and in some cases the responsible programme officers were travelling or had left the organization, and were unavailable in the scheduled time-period (the reasons are analysed in the method section of the report; see annex 3).

With regard to the latter category, could there be any systematic difference between those that were reached and those that were not? A cautionary reply would be: yes, probably in some cases. The fact that some programme officers were unable to fit the interviews into their schedules at any time in almost one month would indicate that they gave the study – and thus possibly the evaluation itself – low priority. Those that left the organization may have done so because they were generally unhappy about working conditions in the wake of Sida's reorganization, or possibly they were pushed out because they did not fit in for some reason or another. There are thus reason to believe that this group would, at least to some extent, have had something different to say about their experiences with the evaluations, and would possibly have given a more negative bias to the responses. Without any systematic investigation we cannot know this, but the results presented below should be reviewed with this in mind.

The interviews followed a structured format (interview guidelines are enclosed as annex 4). The format was designed by UTV, but reviewed and focused by the authors, Jerker Carlsson and Kim Forss. Each interview was conducted by one of three interviewers: Tove Strömberg, Karin Metell and Lisa Segnestam. The latter two were at the time of the interviews working for Sida at UTV, but Tove Strömberg is an independent consultant. Each interview was summed up in a protocol, which is available for inspection at UTV. Printouts of the interviews were handed over to the authors for analysis. The results are presented in a separate study (annex 3).

There are several weaknesses in such a procedure: the interviews were not conducted in exactly the same way, and the authors missed out on some of the more subtle nuances in the respondents replies that are necessarily lost in a standardized format. We tried to compensate for these inherent weaknesses by arranging two feedback meetings, to share experiences and compare notes during the interview process. Also, one of the authors sat along in one interview with each researcher, to make sure that the process was fairly similar and that the interviews held a high and even standard in terms of reliability. The respondents later had the opportunity to comment on our synthesis of the researchers' notes, and they have confirmed that our presentation follows their findings without any bias.

Another source of uncertainty regards the possibility of real or perceived bias on the part of the interviewers. Two of the interviewers were employed at Sida/UTV, which, among other responsibilities, advises programme officers at the departments in the handling of evaluations. Given the role of these interviewers within Sida, there is a risk that respondents answering to them would abstain from voicing critical views of the UTV. There is also the risk that the interviewers interpret responses in the light of their own bias and prior knowledge of the subject. In short, it cannot be excluded that respondents in some cases may have described their experience with the evaluation in a more positive light than is warranted – or that the UTV interviewers may have been less critically inclined towards such presentations than the independent interviewer.

However, although we cannot exclude such forms of bias, we note that there is no significant difference between the responses recorded by the independent interviewer and those employed by Sida/UTV, nor can the authors trace any indication of bias in the transcripts of the interviews. Most of the questions are descriptive rather than evaluative, and thus the risk of such bias is small.

Given the limits of format and time, it is possible that the interviewers did not have sufficient opportunity to pursue intricate issues of purpose and informal decision making.

Quality assessment

The quality of an evaluation system is of course important. But what does quality mean in the context of evaluations? As a starting point, we suggest that Sida refers to the Programme Evaluation Standards, approved by the American National Standards Institute (ASNI) in 1994 and in wide use today in professional contexts as a guide to what constitutes a reasonable evaluation. These were developed over several years by the so-called Joint Committee on Standards, established by the American Evaluation Association, in collaboration with a large body of professional associations, consultants, academics and public authorities. It would also be natural to refer to Sida's own quality standards, as indicated in the Policy, or to the quality criteria established by the Development Assistance Committee of the Organization for Economic Cooperation and Development (OECD/DAC). However, for the sake of simplicity and comparability to other studies, the tool chosen for this study is ASNI's Programme Evaluation Standards. In our view, these are both more general and more precise, and thus provide a better guide to the concept of quality.

ASNI's Programme Evaluation Standards defines and sets criteria for four necessary and sufficient attributes of quality: utility, feasibility, propriety and accuracy. Evaluations that meet ASNI's criteria for all four may be deemed to be sound and fair. It should be noted that the standards relate to each other. An evaluation that is not feasible is not likely to yield accurate conclusions, and conclusions that are not accurate are not likely to be used. Conversely, an evaluation that is conducted according to high standards of propriety will generally have much higher utility than one with shortcomings in this regard.

In this study, the emphasis is on standards for accuracy. These establish whether an evaluation has produced sound information. The evaluation of an intervention must be comprehensive, that is, the evaluators should have considered as many of the programme's identifiable features as was practical, and should have gathered data on those particular features judged important for assessing the project's worth or merit. Moreover, the information must be technically adequate, and the judgements made must be linked logically to the data.

In a sequence of independent research reports, the authors have used an instrument that was originally designed to assess the quality of the evaluation system of the US Agency for International Development (USAID) (Watson, 1988). The same instrument was also used by Canadian International Development Agency (CIDA) to assess the quality of its evaluations. This lead to a synthesis report comparing the two North American donor agencies (CIDA, 1990). A few years later the same format was introduced in Scandinavia and applied in a study of learning and evaluation in the Norwegian aid administration (Samset) (Forss and Haughn, 1991).The format has later been used in Sweden in a study of 177 evaluations (Forss, 1994). The database used in that study was inherited and further developed by us, and it has now been expanded to include 314 separate evaluation studies. These are analysed and the findings presented in regularly appearing research reports (Forss and Carlsson, 1997; Carlsson and Forss, 1997; Grojer, Carlsson and Forss, 1997). The model has also been applied to the study of quality in other aid agencies (Forss and Laursen, 1997).

The data sheet, in its present form, is enclosed in annex 6. It has been changed over the years, but the basic elements are much the same as in the original USAID study. The format has several drawbacks. For one thing, it is difficult to apply without extensive training and dialogue with persons with previous experience in using it. The data entered on it is both of descriptive and evaluative nature. It is the latter that may create difficulties. One third of the data fields regard descriptive variables. sector, cost, composition of team, whether there is a summary or not, whether the report is written in English, etc. The remaining two thirds regard information that requires an evaluation. the person completing the form is required to judge whether the analysis made in the evaluation under scrutiny is "excellent", "adequate", "minimal" or "non existent". This applies for example to calculations of efficiency, the design and application of a questionnaire or the analysis of gender impact. The judgement has to be done with a fair and open mind, keeping in mind how the issues were valued in previous reports.

But if it is possible to cope with these difficulties, there are several benefits of using this instrument. The main benefit is that it is possible to compare the results with those of other studies. At the same time it is necessary to remember that the quality of an evaluation has other dimensions than those mentioned in the data sheet. We provide a starting point for an analysis of these, but our efforts stop when we leave the quality of an evaluation report *per se*, and instead look at the evaluation as a process, where quality is properly measured using criteria that relate more to the nature of the work itself, the levels of participation and the resources available. It is therefore important to correlate results of the quality assessment with information gleaned from the interviews

The format for assessment has to some extent been modified by Sida/UTV, and is thus not identical to the format as used by the authors in other circumstances. The changes are mainly in the descriptive area, where Sida deemed it was desirable to use other classifications of the objects. It should be stressed that these modifications were made by Sida, without any involvement from the authors.

The assessments were made by Karin Metell, Lisa Segnestam and Tove Stromberg. In particular Tove Stromberg has ample experience of assessing evaluations using this format, having assisted us with input to the database on almost 100 evaluations over the past few years. Karin Metell has used the format in the assessment of the first evaluations for the above-mentioned SASDA report (Forss, 1994). Lisa Segnestam was new to the task, but worked in close cooperation with the other two. The quality assessment is presented in a separate report (annex 5), written by Tove Stromberg and reviewed by Karin Metell and Lisa Segnestam There are no differences of opinion on the treatment of the data.

Summary

The present study involves a complex research process, with many persons involved at different stages. The data was gathered in three separate processes, which are described in three separate reports, enclosed as annexes 2, 3 and 5. The text that follows builds on these annexes, and summarizes the information contained in them, but it also analyses the data, and interprets and discusses the issues around the use of the evaluation instrument.

It should be noted that we as authors were not involved in the overall design of the research process, nor with the initiative behind the study. Our roles are confined to two isolated tasks: to write a report summarizing the information gleaned from the interviews with the programme officers, and to synthesize the three reports into a main text. We were not consulted in choice of methodology, the selection of respondents, the design of the questionnaires (apart from a small input, at a very late stage, as noted above) or in the changes in the format for quality assessment.

We wish to emphasize that this study is "owned" by UTV. As conceived, both the compilation of data and the analysis was to be done by UTV staff. Given time constraints, however, it became necessary for UTV to recruit outside assistance in order to time production within the greater context of ongoing efforts in the theme area of change and learning at Sida. As we address similar issues in our research on evaluation, we were in a position rapidly to assess the data.

As ours was but a limited task within the overall research process, our involvement should thus not compromise UTV's overall ownership of the study. In particular, we want to stress that our texts are purely descriptive and analytical. We draw conclusions of an evaluative nature on where the strengths and weaknesses lie in the use of the evaluation instrument, but we do not answer the question on how to improve the situation. That is not our task. Therefore, this report has no recommendations on the future direction of evaluation work at Sida. We assume that UTV will elaborate such practically oriented conclusions with these findings in mind.

2. The organization of the evaluation function at Sida

The Department for Evaluation and Internal Audit

Possibly the most interesting feature of the evaluation system at Sida lies in the combination of a widespread – even diffused – operational responsibility at the level of departments, and a centralized, semi-independent evaluation function. This chapter describes the evaluation function and analyses how it works against the background of interviews with evaluation coordinators. (Note that in this report we only discuss the evaluation activities of UTV, not the internal audit function).

The Evaluation Policy of Sida states:

> Evaluations ... are implemented by the sector and regional departments and by the Department for Evaluation and Internal Audit (UTV). The latter also has an overall responsibility for all evaluation activities in Sida.

In many other aid agencies, evaluations are conducted only by an evaluation unit, which is usually not in a line position (as for example in the British ODA, in the European Commission, and formerly in the Ministry of Foreign Affairs in Norway). As a consequence, evaluations are often not "owned" by people with functional responsibilities, and there is a problem in putting the evaluation findings to use. On the other hand, a system where all evaluations are formed in response to the problems perceived by management can loose its credibility. Evaluations have both an external and internal audience, and the design at Sida reflects an ambition to make evaluations useful for several stakeholder groups.

We called the UTV's status "semi-independent". Why not simply call it independent? UTV is in fact more independent than many other evaluation units that do call themselves independent. UTV reports to Sida's Board of Directors on a regular basis. The Board has adopted the Evaluation Policy, and it approves the annual evaluation plan that is presented by UTV. However, it is the Director-General who allocates the budget for the department, and the personnel are all employees of Sida. The Director-General is also Chairman of the Board. UTV's follow-up reports to the Board encompass all evaluation activities at Sida, and the unit operates in close symbiosis with the rest of the organization, for example as regards personnel, budget and career paths. The word "independent" is much misused in the context of evaluations, and in our opinion the term should be confined to situations where there is no mixture of responsibilities, where there is no dependence on a budget allocation, and where the reporting is to a body which clearly has no vested interest in the outcome of the function.

The responsibilities and activities of UTV with regard to evaluation can be described as falling in three categories.

In the first category fall the analytical and administrative tasks of conducting evaluations, or, to cite the policy document, "to initiate, design and perform comprehensive and for Swedish development assistance strategically important evaluations of Sida activities". Under these terms, UTV is tasked with initiating and carrying out evaluations under its own auspices or in cooperation with other organizations.

In the second category fall functions relating to the evaluation activities of other departments. The task is, to cite the policy document, to disseminate information and recommendations from evaluations, in cooperation with the operative departments, with the aim of improving the analysis, assessment, design and implementation of ongoing and new development assistance activities, to propose changes in development inputs, organization and operational forms as a result of evaluation activities and with the above mentioned aim, to supervise, examine for quality and improve Sida's evaluation system generally.[1]

These overall responsibilities are operationalized with the following tasks:

> [T]o coordinate the work of producing an annual evaluation plan and report for Sida, . . . to assist the evaluation system with advisory services focusing on quality assessment of terms of reference and evaluation reports and on choice of methodology and of consultants, development and maintenance of the report and information database; development of methods and concepts of importance to evaluation activities, design and dissemination of rules and guidelines for evaluation activities including documentation and reporting, to participate in the work of Sida's project review committee.

The third category relates to international cooperation. UTV has the responsibility "to participate in international joint evaluation work." This involves efforts

> to exchange experience and cooperate with the evaluation functions of other development assistance and international organizations, and to support and contribute expertise to the development of evaluation capacity in recipient partner countries [2]

The evaluation policy does not say anything about the balance of work between these three categories, but it is notable that the functions are very different. A week's work spent in one area of activity will have little significance for the two others. The audiences are different, and the output cannot be measured by the same standards. Conducting evaluations under one's own auspices is an activity which is relatively easy to control, and which yields a clear and identifiable product. Supporting other departments with input on, for example, the formulation of terms of reference or the choice of methodology, is far more uncertain and difficult, and perhaps less interesting from a professional point of view.

Can it be assumed that the three categories are equally important? Does the amount of text in the Evaluation Policy reflect their relative importance? Of a total of 220 words that describe tasks and responsibilities, 30 words are used for the first category, 152 for the second category and 38 for the third category. This could be taken to indicate that the advisory services to the line departments are the most important tasks. As the bulk of evaluation studies are initiated and used at the department level – or even at the level of divisions within the departments – such a priority would presumably have the highest impact. On the other hand, the processes themselves are not controlled by UTV. We cannot answer the question on relative importance of UTV's three tasks, but the issue should be kept in mind when reviewing the analysis of the evaluation process presented in chapter 3.

Evaluation work in the line functions

According to Sida's evaluation policy, as part of its responsibility for operational activities and results, each of the 14 operative departments are to initiate, plan and implement evaluations of programmes, projects, inputs and activities within its area of activity. UTV focuses on more comprehensive and strategically important evaluations. This division of responsibilities presupposes regular consultations

---

[1] *Sida Evaluation Policy* (Sida Stockholm, 1995), p 4.

[2] Sida (note 1)

between UTV and the operative departments, including the embassies in countries with extensive development cooperation.

Each of the departments appoint a coordinator for evaluation purposes. This function was established in 1995, and all the coordinators were selected within the year. The evaluation coordinators, together with UTV, constitute a network for evaluation. At regular meetings with UTV staff, the coordinators represent their departments. Within their departments, they are expected to report on the contacts with UTV and the rest of the network. Practically speaking, the network has the following tasks (protocol from a network meeting on 18 December 1996; these suggestions were discussed, but have so far not been formalized):

1 identify the need for competence development with respect to evaluation and plan training activities;
2 coordinate the evaluation activities of the department and prepare the departments annual evaluation plan;
3 assign priorities in connection to the annual evaluation plan, in cooperation with UTV;
4 report on the accomplishments in evaluation and inform the network on the department's compliance with the evaluation plan; and
5 together with UTV, assist the departments with advice on the evaluation process.

The first evaluation plan for fiscal year 1995/96 called for 117 evaluations to be implemented within 1996. Of these, 81 were initiated in the same year, along with 33 evaluations not mentioned in the plan. There are significant differences both between and within departments on the overall management of evaluations of projects and programmes. At the regional departments, the authority for most projects is delegated to sector departments or embassies. Consequently, these departments seldom commission evaluations. The regional department for Latin America is an exception, as it has more operative responsibilities, and thus also a more active role in coordinating and commissioning evaluations. The departments for Eastern and Central Europe (Sida-Öst) and for Research Cooperation (SAREC) initiate and commission several evaluations every year. But it is the three sector departments (INEC, DESO and NATUR) that account for the lion's share of Sida evaluations.

None of the departments have any specific guidelines for evaluation, nor are there any sector specific guidelines (apart from Sida's Evaluation Policy). The actual procedures appear to have developed out of practice, and to some extent the programme officers do as they always have done, which sometimes means that they do what they did in the organization they worked for before the new Sida was created. Below we analyse the coordinators experiences of how the evaluation instrument has been used in their respective departments, from the first initiative through the completion of a report and to the use of the reports within the departments.

Assessing the experience of the coordinators

As mentioned above, 9 of the coordinators were selected for interviews by UTV. Let us now analyse what their experience of the coordinator function is after one year. We should bear in mind that this is a relatively short period for gaining significant experience. In addition, there has been a considerable turnover, as 7 of the 14 coordinators have transferred, with new ones taking their place. At present, 8 of the 14 coordinators are women.

The interviews provide some background on the coordinators. Of the 14, 4 have worked with development assistance for more than 20 years, 2 for 10–20 years and 3 for less than 10 years. All except one have experience of long-term postings in developing countries. Most have a professional background within former SIDA, with 6 having had their current task for more than 6 years. Most have extensive experience in evaluation work within the agency, having several times initiated, managed and followed up on evaluation processes.

However, they have very limited experience of actually conducting an evaluation. Only one has taken part as an evaluator, and only three have at some stage in their career joined evaluation teams as observers. Their actual experience of what it is like to undertake an evaluation is thus rather weak, and it may be assumed that their knowledge of evaluation methodologies have significant gaps. Their perspective on the task is likely to be the same as that of their colleagues in the department, and that may inhibit the usefulness of their role. To be effective as an advisor one should ideally be able to offer something extra, either substantial theoretical competence or practical experience of the evaluation process.

The coordinators all state that there are no formal rules or procedures governing the choice of evaluation objects. It is common that more costly projects are selected, or those of long duration. It is also common that projects are evaluated in anticipation of an extension review process. In such cases the evaluation may serve as a reference document in negotiations with partners. The selection of evaluation objects is also determined by the extent and nature of problems encountered in the projects. Projects where the progress reports provide satisfactory information, and where the programme officers do not perceive any particularly problematic issues, are not likely candidates for evaluation.

The interviews with coordinators also suggest that the initiative to evaluate usually comes from the programme officer in charge of a particular project or programme. It is rare that evaluations are initiated at higher levels in the organization. Usually people arrive at a consensus, when they review the likely evaluation objects. It is also common that agreements with partner countries contain a standard clause that projects are to be evaluated at a half-way point, at completion or prior to an extension review. During annual renegotiations these clauses may be activated. To date, regional departments are said to have little influence on the initiative to evaluate, despite the fact that evaluations are supposed to be an important input to their work on country strategies.

According to the coordinators, most of the work in preparing an evaluation occurs within the division. The programme officer usually prepares a first draft of the terms of reference. The head of the division may be involved at this point, depending on his or her interests. The formal decision to evaluate is always taken by the head of division, but this is a formality that merely confirms the previous steps . This formal decision is taken in connection with the allocation of budget funds, that is, when a contract is signed with evaluation consultants.

There are few contacts with other parts of the organization – most of the work occurs on the division level. The exceptions occur when there have been transfers of personnel. A new programme officer will normally contact his or her predecessor, even if that person has moved to another division. An advisory service that spans over several divisions has a limited impact, and so does the policy department, as well as UTV. According to the coordinators, UTV seldom has any influence on the formulation of terms of reference or on the choice of evaluator – usually, the department is simply not involved. Furthermore, according to the coordinators it is common practice to exchange the draft terms of reference with partners in the recipient countries and to ask for their comments. Nevertheless, the real influence of the recipient is assumed to be minimal, except with regard to the Department for Research Cooperation.

In conclusion, the interviews with the coordinators imply that the programme officers' engagement in evaluation work is a lonely task, but one integrated with his/her current duties. Personal contacts, mainly within the division, determine the extent of consultation. There are no formal requirements regarding consultations, and the institutional mechanisms for accessing the advisory services at UTV and the evaluation network (of which the coordinators are part) are not much utilized.

How do the coordinators' assess the tendering processes and the selection of evaluators? Again, within the wide framework of governmental and agency-specific rules on purchases and contracts, there are no specific regulations or guidelines. But it is common practice to select consultants that are independent from the project or programme . He or she must not be biased. All the coordinators emphasize this aspect of qualification.

In general, the coordinators agree that it should be possible to engage Sida staff in evaluations, but not those who are directly responsible for the project. In practice it is rare that Sida personnel are engaged as evaluators. The coordinators are also sceptical of having evaluators from the recipient countries, but here the departments differ. Those with a background from the former BITS follow a practice of engaging Swedish consultants as evaluators, while those from former SAREC are said to actively encourage the use of consultants from developing countries. Other departments usually appoint teams, and they strive to have a team member from the recipient country as well, in the hope that this would bring knowledge of local conditions that is necessary in the evaluation.

The appointment of evaluators is said to vary between departments. The sector department for infrastructure and economic cooperation (INEC), and Sida-Öst are said to use competitive bidding more frequently than other departments (however, as is shown below, this is not true). Other departments' coordinators recognize that evaluations seldom are put on tender None of the departments have any roster for consultants or evaluators. A few divisions maintain such rosters, but they lack search keys that would allow accessing people with evaluation competence. When they exist, it is rare that consultancy rosters are used.

It is unclear how programme officers select consultants for evaluation, but the coordinators suspect that personal contacts and recommendations from colleagues play a major role. Personal reference from colleagues clearly carry much weight in the choice of evaluators. A common practice would be to select consultants who have successfully completed a similar task once before. The responses suggest that there are considerable weaknesses in the selection process. There is a large risk that the same consultants are used over and over again.

The coordinators do not generally know how the costs of an evaluation are calculated within their departments. They assume that an estimate of the costs is calculated once the draft terms of reference are formulated. By then it has become possible to form an opinion of the magnitude of work involved. The terms may then contain an indication that, for example, 4 weeks may be used for the evaluation. The tendering documents, to the extent there are any, would normally suggest what would be considered a reasonable work load. There are no rules, and no explicit recommendation on how the costs are to be calculated, or how much an evaluation may cost. (Sida has recommendations on the level of consultancy fees, and a ceiling on how high fees that can be accepted). But according to the interviews, the cost calculation is often limited to an implicit assessment of a reasonable level of expenditure, based on the questions to be answered, the total cost of the project and the expected usefulness of the evaluation (!).

There are many ways to finance an evaluation. If the evaluation was agreed with the partners when the project was formulated, it is possible to finance it from the country frame, or there may even be funds set aside within the project budget. It is also possible to use the so called IRV budget (funds for project assessment and evaluation). The coordinators conclude that it is not a problem to find financial resources to conduct evaluations.

Once the evaluators produce a draft report, the programme officers submit copies of the draft to the concerned stakeholders: project consultants, line ministries in the recipient countries and the authorities and organizations involved in implementing the project. The comments from the stakeholders are usually forwarded to the evaluators, who then produce a final report. The dissemination of the completed reports depends largely on the programme officer, and how he or she assesses the interest and demand for the report. It is rare that project evaluations reach any but those directly concerned with the activities. None of the departments have systematic procedures for the dissemination of reports, whether internally or externally. According to the coordinators, two departments – INEC and NATUR – organize seminars to make the evaluation results known.

None of the departments are said to have either rules, routines or any informal procedures to assure that evaluation results are used. Although the draft report usually contains recommendations. there is no requirement that programme officers or division heads take a position on these. None of the coordinators knew of any follow-up on evaluation recommendations, either at the level of divisions or departments.

The role of the coordinator is thus rather difficult, as is the role of UTV. Evaluation in the functional departments appears to be a very decentralized process, left to the initiative and professional competence of the programme officers, and undertaken in little contact with the rest of the organization. In the absence of routines, policies and procedures at their level of work, that is, within the department, most of the work seems to be of an arbitrary and haphazard nature.

Not surprisingly, the coordinators report limited impact of their function. They do disseminate information from UTV to their departments, and they do also assemble information from the department to UTV. But the range of their activities is open to doubt, and it can mean anything from telling one or two persons about UTV activities to more comprehensive meetings, open letters, or leaflets.

The coordinators believe that their function is known within the departments, but only one of them had taken part in the department's planning of evaluations. But the coordinators had taken part in the semi-annual and annual follow-up meetings on the evaluation plan. The coordinators say that their roles with regard to their respective departments' evaluation activities are unclear, and their colleagues appear not to understand it fully. 5 of the 9 coordinators say that senior management has a "positive" attitude to their work, although this does not show in practical work. The other coordinators report "neutral" attitudes to their role. None of them had any changes in job descriptions nor in the range of their duties as a consequence of becoming evaluation coordinators. 4 of the 9 find the task stimulating, but there are also 4 who found the task a heavy burden. Most of them still think that their role as coordinators is necessary – but also that the function would not have been created if UTV had not asked for it.

In sum, the coordinators have an understanding of the evaluation process as something very decentralized, highly personal and informal, relevant only for those most directly concerned with the particular project or programme being evaluated, and much dependent on the context of each particular evaluation. Their own role is uncertain and not much in demand. They encounter little encouragement and understanding among management and colleagues, but not much opposition either. In fact, it appears as if nobody cares – to any visible extent – about the efforts at streamlining, coordinating and improving the quality of evaluation work.

As for the role of UTV and the three categories of work mentioned above (conducting strategic evaluations, supporting evaluation in operative departments and international cooperation), the interviews imply that the actual practice in conducting evaluations has not changed much – but that there may be a need to pay far more attention to it.

3. The evaluation process

Introduction

In this chapter we look closer at how the evaluations that were conducted during fiscal year 1995/96 were managed. The analysis is based on information gleaned from interviews with 30 programme officers who were in charge of one (or several) of these evaluations. The format of the interviews is found in annex 4, and in annex 3 there is a more detailed description of the responses. This chapter builds on the descriptive material in annex 3, supplemented by information coming from the interviews with evaluation coordinators, as well as with information from the quality analysis. While annex 3 provides a descriptive discussion of the results of the interviews, this chapter presents an interpretation and analytical assessment of the material.

At first, we need to look at the background of the respondents. First of all, which sectors were the evaluations found in, and in which divisions do the personnel work? Of the 30 evaluations, 7 were produced by INEC, 7 by DESO, 7 by Sida Ost, 5 by SAREC and 4 by NATUR. Many of the respondents have worked a few years in several of the aid agencies. Those that have worked longest in aid are normally found at SIDA, but others have worked at BITS and the present Sida, or at SAREC, SIDA and again at Sida. Some have worked for private consulting firms, others for UN organizations or for NGOs. Some have a background in development research. Somewhat more than half (18) have experience of living in developing countries for more than a year, whereas the remaining (12) have never been in these countries except on short visits.

The interviewers ask how frequently the respondents are involved in evaluation tasks. It is clear that evaluations are a major ingredient in the work of most respondents. All had experience in formulating terms of reference and receiving and following up on evaluations. Only a few had never initiated an evaluation of a project or programme, and only one had never contracted evaluators.

But even though most of the respondents work with evaluations, few have much experience in conducting evaluations themselves. Those who have usually have a background in consulting firms specializing in development and aid, and we assume that these experiences date back to this previous employment.

In terms of background, career patterns and involvement in evaluation work, there are thus obvious similarities between the 30 programme officers and the 9 evaluation coordinators who provided the data for the previous chapter.

Why are evaluations started?

It is often said that Sida is a very decentralized organization, and this seems to be confirmed with regard to most decisions concerning evaluations. In 17 cases the respondents were actually the ones who initiated the evaluation. In fact, it is generally the programme officers themselves who initiate the evaluations. It is rare that an external agent, or any other part of the organization, takes the lead Exactly the same opinions were forwarded by the evaluation coordinators.

Needless to say, the formal decision to evaluate is always taken by the head of the division, and thus he or she also gets involved at an early stage. The most frequent response to the question of how the evaluation was initiated was: "I came up with the suggestion and then discussed it with the head of the division, who took the formal decision".

It should be noted that the formal decision comes rather late in the process. The head of the division agrees to prepare the evaluation exercise at an early stage, but a formal decision to evaluate is not taken until money is allocated, and that usually does not happen until the terms of reference are set and a budget had been agreed upon with an evaluation team. In other words, the formal decision only confirms a process which by then is already well under way. In one instance the respondent says that the informal decision to evaluate was taken in May 1995, while the formal decision was taken in March 1996. The question then arises why the decision to initiate an evaluation of a particular programme or project was taken at all. Here there are a number of different responses.

The most common is that an evaluation has been mentioned earlier in the management of the project. In some cases it is called for in the project document, or in the agreement with other stakeholders. In some divisions it is simply common practice to evaluate a project after a given period of time. In 9 instances this is mentioned as the reason for initiating the evaluation. On the other hand, in the interviews the evaluation coordinators clearly state that there are no sector-wide or departmental policies on when and how to evaluate, so the "common practice" which these respondents refer to must be implicit or even personal.

Some of the respondents mention that the evaluation object was chosen in order to have something to include in the evaluation plan. The activity selected either seemed "suitable" or the programme officer thought it would be "interesting" to have this particular activity included in the evaluation plan. In this connection, it was often mentioned that the selected activity had never been evaluated before, and that it thus made good sense to include it in the evaluation plan. These slightly overlapping reasons were also mentioned in 9 instances, where the need for something to include in the evaluation plan was cited by 4 respondents and the wish "to know about results" was cited by 5 respondents.

Another reason given was that the project had come to an end, or was approaching a mid-term review. In both cases, Sida would soon be entering negotiations on whether and how the programme should continue. Here an evaluation would contribute to the decision-making process. This is mentioned as the reasons to evaluate by 4 respondents.

3 evaluations were initiated because the division in question was expected to make a contribution to the development of country strategies. The recent reorganization of the aid agencies has also influenced the choice of evaluation objects. In two cases the evaluations were initiated to document former BITS projects and that organization's methods of working with aid.

Several interviews do not give any clear response to the question, and often the responses seem to overlap. An evaluation was perhaps initiated to generate information needed for the work on a country strategy *and* because it was required by agreement with other stakeholders. There could of course be other combinations of motives.

However, the initiative to evaluate was usually not taken in isolation. To what extent were other stakeholders consulted when the decision to evaluate was taken? Responses are presented in table 3.1. Note that in theory each category of stakeholder could be represented 30 times, but in practice the range of contacts is far lower. The most frequent contacts are with the programme officers' immediate superiors: heads of the divisions or subdivisions. A total of 82 stakeholder consultations are mentioned in the interviews, which means on average roughly three per evaluation.

Table 3.1 Consultations with other stakeholders when the decision to evaluate is taken

Stakeholder category	Number of respondents indicating that they had contact with someone in this stakeholder category
Evaluation coordinator	1
Head of division/department	24
Colleagues in the division	9
Evaluation secretariat	5
Other units at Sida	4
Embassies	9
Recipient country organization	12
Project consultants	9
Other consultants	6
Other donor agencies	3

Decision-making processes are complex phenomena. When charting them, it is important not to interpret the real-life situation too mechanically, nor to expect the process to follow a stringently rational course, like a text-book example of management theory. The closer the model gets to reality, the faster do such theories break down. But the challenge to understand what happens remains, and so does the quest for good performance. Let us first analyse the purpose more closely. What do the respondents have to say about why their respective evaluations were started?

The responses to this question are comprehensive and explicit. In fact, many respondents quote the objectives stated in the terms of reference for the evaluation. Below follow some examples of what the programme officers said with regard to different objectives.

First, and most important, is immediate managerial use. All of 17 respondents mentioned managerial use as the most important reasons for the evaluation, as the following responses serve to illustrate.

> We intended to use the evaluation results in future decisions on whether (this consulting firm) could be used in the future

> The purpose was to assess the programme and give recommendations for future support This could be used in planning, rectify problems, determine how much money would be needed, and improve the planning.

> We were uncertain about the effects of the programme, and what results that had been achieved The project had lasted for several years, and there was now a proposal to continue, and thus we needed an evaluation

If we classify the responses into three broad categories of evaluative purposes – audit, management and learning – we find that an overwhelming majority of the respondents identify management needs as the prime motor for the evaluations, followed by learning needs. Only very few evaluations were initiated to satisfy audit needs. At the same time, we are left with the impression that evaluations are initiated on loose grounds – because it is felt an evaluation is overdue, because something has to be put into the evaluation plan or because it has been suggested in the agreement with the recipient country.

There is nothing wrong in this *per se* – it lies in the nature of an initiative to be vague. But the elaboration into a more full-fledged purpose does not seem to occur. The first vague initiative is rapidly followed by fairly standardized terms of reference, often written with the assistance of the evaluation manual. Indeed, rather than elaborating on the initial reason, thinking through what information is needed and setting priorities among conflicting purposes, the terms of reference are almost copied from the manual. There is a lack of practical focus at the level of the individual project or programme, and instead we find shopping lists that are called purpose. Perhaps the use of standardized terms of reference make it too easy for the programme officers to define the evaluation task?

How are terms of reference defined?

It should be kept in mind that the terms of reference usually are formulated after consultations with others at Sida as well as with outside stakeholders. The respondents were asked how widespread such contacts were. The results are presented in table 3.2 (note that one respondent did not know who was contacted, so the number of responses is only 29).

Table 3.2 Consultations during the preparations of terms of reference

Stakeholder category	Number of respondents indicating a contact
Evaluation coordinator	0
Head of division/department	23
Colleagues in the division	15
Evaluation secretariat	12
Other units at Sida	9
Recipient country organization	12
Project consultants	0
The evaluation team	6
Other donor agencies	7
Other Swedish agencies	2

It is clear that the contacts are still few and far between. The most frequent contacts are naturally with the programme officer's unit or division head. But perhaps surprisingly, not all respondents had such contact with their immediate superior.

By and large, there are more contacts within Sida in preparing the terms of reference than in initiating the evaluation idea. But only 12 of 30 respondents had any contacts with the recipient government organizations. There were also few that sought contact with other donors, or other organizations.

The evaluation coordinators played no role at this stage, even though assisting the programme officers in the formulation of the terms of reference is intended to be one of their primary tasks. Table 3.2 also shows that the coordinators tend to overestimate the range of consultations, particularly as they thought it was common practice to consult with the recipient country authorities. It would be interesting to see whether evaluations that were preceded by extensive consultations with recipients and other stakeholders tend to be of better quality. Our data does not allow a full-scale regression analysis, and of the three evaluations in the total population that was rated "excellent", only one was followed up with an interview. In this particular case, the range of contacts is limited to the head of the sector division, the heads of two other sector divisions and the three organizations that were implementing the project. This would suggest that more extensive consultations do not generally lead to better evaluations, but we should not draw too extensive conclusions from one case.

Another question of whether consultations with other stakeholders influenced the terms of reference or in any other way resulted in changes in the direction of the evaluation. One of the evaluations is reported to have been changed considerably in the light of consultations, 18 were not changed, and for the remaining 11 there were minor changes. These were of two kinds: it was suggested either that some specific item should be added – notably a discussion of project/programme cost-effectiveness – or that the evaluation should be better focused or better structured. The former comments mainly came from UTV, the latter often arose out of contacts with division heads or with other units at Sida.

If admonitions like "don't forget to assess cost-effectiveness" are what typify UTV's input at this stage (which we have no way of knowing), it is perhaps not surprising that the department is not consulted more often. Most of the programme officers have read the evaluation manual, which contains draft standard terms of reference, and one of the items to be covered is cost-effectiveness. The comment as such cannot be said to be useful unless it also involves practical guidance on how to go about the task.

Careful preparation of the terms of reference does require some sort of research. The programme officers interviewed here have used project documents, quarterly reports, annual reports, other progress reports or special reports from consultants, previous evaluations (if any) and the original agreements with the recipient country. All have consulted these sources. A majority had read, and referred to, Sida's evaluation manual. They were asked whether they knew about Sida's evaluation policy, and most did, but many were not aware of it at the time of writing the terms of reference.

A majority are satisfied that they had found the documents needed, most of which were available at their respective divisions. The archives were not used. Four respondents mention problems in finding background documents, and one could not answer whether the necessary documents were found. In one case the respondent explains that as the evaluation consultants wrote the terms of reference, the respondent did not know whether it was difficult to find the documents.

How are evaluations budgeted?

Judging by the data, the 30 evaluations cannot be described as expensive. Costs range between approximately SEK 200,000 and SEK 600,000. Nevertheless, in aggregate this means an expenditure of some SEK 20 million for evaluation activities at the level of the operative departments. It should be of some concern how that money is spent, for tax payers as well as for Sida's management.

24 of the respondents said that they made a rough estimate based on the approximate number of working weeks that would be needed, the extent of travel and a ball-park figure for the fee usually charged by an outside consultant. Two respondents state that there was a ceiling for evaluation costs, but they do not know how that ceiling was calculated. Two respondents state that the evaluation was put on tender and that the consultants had to suggest a budget. Price then became one of the selection criteria. In both cases no budget estimates had been made prior to the tendering process. Finally, two respondent did not know how the costs were calculated.

The critical question here is how the respondents arrived at the estimate that, say, 5 weeks of work would be needed. Is that estimate based on sound knowledge of what it requires to gather data to substantiate a calculation of effectiveness or efficiency? Is it based on a clear understanding of how one would arrive at reliable conclusions with respect to sustainability and impact? Are there any indications of such methodological awareness either in the terms of reference or in the evaluation reports themselves?

It is reasonable to assume that the costs are not based on careful analyses of evaluation methods. But what about the evaluators themselves, would they not know whether it is possible to answer the questions in the terms of reference within the number of weeks allocated? Maybe, but on the other hand they may assume that the client will understand what type of answer he or she will get for the money. Some sort of an answer can of course always be provided.

The evaluations were financed from the IRV budget (10 evaluations), the country frames (7), the budget heading for contract-based technical cooperation (7), a separate budget line for environmental studies (2) and in one case from a regional budget line. None of the respondents indicated that the evaluations were insufficiently financed. One even felt that the evaluation was over financed. (In 3 cases we could not determine how the evaluation had been financed.)

Is there something in the recruitment of evaluators that tend to make the evaluations more costly than they need be? 25 of the 30 evaluators/evaluation teams were contracted directly, without any contacts with other "competitors". Naturally, this does not give the evaluators any incentives to suggest low-cost methods, or to find cost-saving alternatives to rather expensive personal travelling and personal interviews, in particular as the number of working weeks appears to be specified from start.

The most common reasons given by the respondents for why not more evaluations were put on tender were that the budgets were too small, there were so few persons that qualified, or a tender procedure was administratively cumbersome. Sida has rules to determine when programme officers have to hire consultants after competitive bidding. The interviews suggest that these rules are not well understood. Some mention a cost ceiling of SEK 200,000, others a cost ceiling of SEK 300,000, above which a task must be put on tender. Another issue highlighted in the interviews is the difficulty of assuring proper observance: one respondent describes how he/she side-stepped the rule by splitting the total cost in two, placing fees in one basket and reimbursable items in the other. It was thus possible to avoid the competitive bidding procedure.

While the evaluation coordinators seem to be aware that tendering is not so common, their replies give the impression that it is more frequent than it actually is. Only 5 evaluations of the 30 covered by the interviews were put out to tender with competitive bidding, which is 17% in one year. Interestingly, the coordinators believe that INEC and Sida-Ost use more competitive bidding for evaluation than other departments. But of the 5 that were put on tender, 1 was commissioned by DESO, 2 by NATUR and 2 by INEC (which translates into 20, 50 and 25 per cent of the respective departments' share of evaluations in our sample).

At this stage, it would be interesting to compare if there are any indications that evaluations put on tender are more professionally conducted than others. The only one among these 30 that was judged "excellent" in the quality analysis was one of the five commissioned through competitive bidding. Of the remaining, two were in the group judged to be "good", and the other two were "adequate". The sample is small and does not justify any general conclusion, but we can venture a hypothesis that evaluations commissioned under competitive bidding stand a fair chance of being of higher quality than evaluations that are not.

In light of the fact that most of the respondents state that it was not difficult to find suitable consultants, and that there were many to choose from, it is surprising that competitive bidding was not used to a greater extent.

How are evaluators identified?

The evaluation teams, or the evaluators, were mainly identified through informal personal contacts. The programme officers asked colleagues in the division, people at the embassy or even outside organizations. Many of the selected evaluators were already well-known to the respondents (15 cases). Few respondents made any significant effort to locate new evaluation expertise. Only one turned to the consultancy roster at UTV to find names. (Which is surprising, as UTV claims it has no such roster!) The fact that programme officers tend to stick to the same, rather small, network, without trying to locate new sources of expertise, makes it difficult to ascertain how hard it is to find other evaluators. Perhaps it is difficult, perhaps it is easy.

The type of competence requested by the respondents varies, but most mention sector competence and knowledge of specific subjects as prime considerations. Many also looked for someone with a knowledge of the country, the region and the local language. Other valued qualifications are evaluation experience, knowledge of development issues and knowledge of cross-cutting issues (gender, the environment, democracy and human rights, free market reforms, etc). Skills in financial analysis are called for the least, and are mentioned by only two respondents.

Of the 30 evaluation teams, 6 include local consultants. A few respondents indicate that they had wished to include local evaluators but that for various reasons this was not possible. Of the remaining 20 respondents, many never considered the idea and most were sceptical, arguing that it would have been impossible to find someone with the necessary skills.

How does this correlate with our assessment of the best evaluations? The only evaluation rated "excellent" had in fact recruited a non-Swedish team member. However, as it concerned a programme with a global scope, there was no identifiable recipient country. In other cases where the evaluation team included a participant from the recipient country , several are rated "good", a few "adequate" and none "inadequate".

The composition of the evaluation teams does not indicate any higher degree of gender awareness. Evaluators are predominantly men. Women are rarely included in the evaluation teams unless the project under scrutiny concerns issues such as women in development, child health or family planning.

We have now analysed what happens from the first initiative to the start of the real evaluation work, that is, when an evaluation team has been recruited and is ready to begin field work. At this stage the role of Sida's programme officers tends to be minimal, as few follow the evaluation teams as observers or participants. The next phase of active involvement comes when draft reports are presented to the programme officers.

It is possible to organize some form of a steering or reference group to monitor the evaluation process. Again, such a group would normally be most actively involved when the terms of reference are formulated and evaluators selected, and again when the draft report is presented. In only 2 of the 30 interviews did the respondents confirm that reference or a steering group had been set up. 2 others state that such groups were planned but were for different reasons never set up. 5

respondents mention that they worked closely with colleagues or with their unit heads, and so in effect formed informal reference groups.

How much time does an evaluation take?

Managing evaluations is just one of many duties that the programme officer is tasked with. It is difficult to say how important evaluation management is for programme officers, but there are indications that it is not among those given top priority. Activities relating to the planning, preparation and monitoring of projects, administrative duties, etc., are probably regarded as more important.[3] Although evaluations are normally contracted to external consultants, programme officers still need to be involved in the process. The question is how much time they actually spend on an evaluation - from start to finish - and what they do exactly when managing an evaluation.

The respondents are not asked to specify how much time they spend on the tasks associated with administrating evaluations. Still, a general description of these activities can shed some light on how time-consuming they are. The planning of the evaluation is closely connected with the formulation of the terms of reference for the evaluation. The nature of this early phase is discussed in detail in sections 3.2 and 3.3 above. Suffice to add that the preparation of the terms of reference involves more than simply writing down a number of questions for the evaluators to answer. It also involves consultations with stakeholders, specifically with the division head, colleagues in the division, the evaluation secretariat and the recipient organization. Whether this is a time-consuming task or not depends very much on the circumstances in each case.

It is during the evaluation of the tenders that the programme officers discuss issues pertaining to the design of the evaluation. Once this has been agreed on, questions regarding methods and theory do not form part of the programme officers involvement in the planning of the evaluation. This does not mean that they are inactive or leave the evaluators on their own. On the contrary, they are very active. The most common participation pattern can be described as "paving the way". One example illustrates the point:

> Karin and the consultant met in Stockholm and discussed the visit in Malaysia. She also participated in the interviews with researchers at the University of Umeå. The consultant got access to all the material which was necessary through Karin The consultant did not cover all areas and Karin therefore asked the consultant to contact an expert on hydrology, who could read and assess those areas in the report Karin also helped the consultant with a letter of introduction before travelling to Malaysia

Programme officers assist evaluators in tracking down background information from agency archives and elsewhere. They also contact concerned parties in the recipient country and in Sweden, inform them of the coming evaluation and ask them to cooperate with the evaluator.

It is rare for programme officers to participate as a member of the evaluation team. The officers resume their involvement when the draft evaluation report is submitted. Again it is difficult to generalize about the time spent at this stage. It

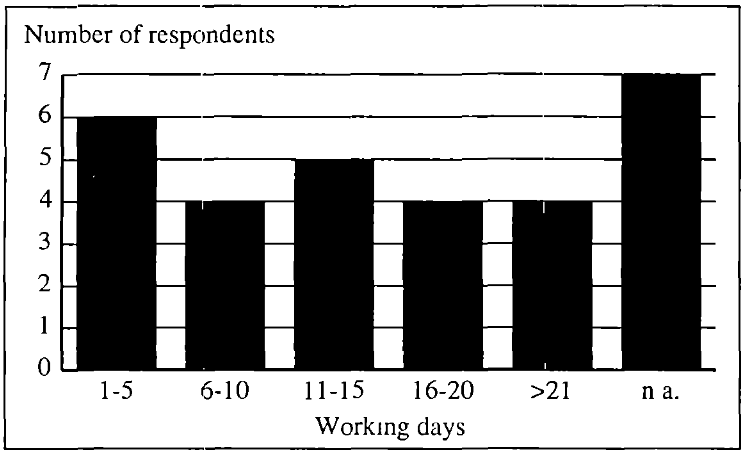
---

[3] See Mosley, P , "The politics of evaluation a comparative study of the World Bank and UK ODA evaluation procedures", *Development and Change*, vol 14, (1983), pp. 593–608; and Mosley, P , Harrigan, J , and Toye, J, *Aid and Power: The World Bank and Policy-Based Lending*. vol 1 (London: Routledge, 1991), pp 45–51.

depends very much on the quality of the report and on the officers themselves. This issues are discussed in more detail in section 3.8 below.

This pattern of involvement usually means that an evaluation exercise can be quite time-consuming. The programme officers were asked to state the number of working days they had spent on their respective evaluations. Figure 3.1 summarizes their responses.

Figure 3.1 Working days spent on an evaluation



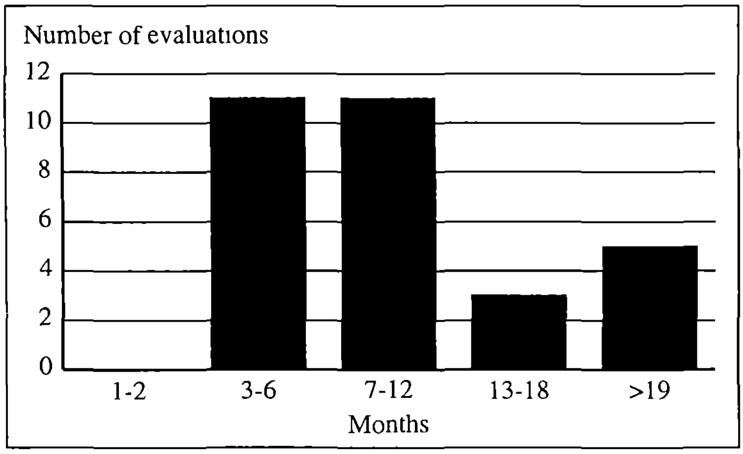
The reader should keep in mind that the figures presented in figure 3.1 are not exact. Most programme officers do not keep time logs. When asked, they simply made a rough estimate. Some of the evaluations date back a couple of years, and in these instances the officers found it difficult to remember how much work they had put in. In some cases the question was not applicable, since they had not been fully involved in the process. With these reservations in mind, the figure still offers some interesting insights.

First of all, although the distribution is fairly even across the range of alternatives, it appears normal for programme officers to spend up to 15 working days on an evaluation. It is, however, noteworthy that as many as 8 respondents spent more than 15 days working on an evaluation. Assuming that an average evaluation involves between 5–7 man-weeks of consultancy work, 3 weeks of a programme officer's time suggests that evaluation management is a time-consuming task.

One may have expected a correlation between the size of the evaluation – in terms of consultancy man-weeks – and the time spent by the programme officers. However, this turned out not to be the case. Large evaluations do not necessarily require more input from Sida staff members than smaller evaluations. The time spent is determined by other factors, such as problems encountered in the field, the efficiency of the consultants, the quality of the draft report, etc.

How long time does an evaluation take? From the point of view of the decision-making process, it is too narrow a perspective to focus only on the time allotted for the evaluation team. One need to look at the whole evaluation process, from initiation to submission of the final report. Figure 3.2 shows the distribution

Figure 3 2 From decision to a completed evaluation



It seems it is difficult to conduct an evaluation in less than three months. In 73 per cent of the cases the whole process took up to a year. In five extraordinary cases it took up to two years. Obviously such lengthy procedures were not intended. Despite careful planning, the evaluations were for various reasons fraught with delays. Some had to do with finalising of the report – language problems and other editorial matters. In some cases it took an unduly long time to have questionnaires filled in and returned. In other instances it took a long time to find all the relevant background material. Delays that occurred in the course of information gathering sometimes upset the evaluators time schedules, which may have provoked further delays later on.

Processing draft reports

The review of the draft report is very important to the different stakeholders in the project under evaluation, as it provides an opportunity to express views on the findings and recommendations of the evaluation and to press for changes. Once the final report is presented, there is little left to do. The draft review is equally important to the evaluator, as it provides feedback on the quality of the analysis as well as on the findings' relevance to the client. In other words, the draft report forms the last opportunity to ensure the quality of the evaluation. One would therefore expect the work around this review phase to be particularly intensive. What happens when the first draft arrives? Who are involved and what do they have to say about the report?

When the first draft arrives, it is circulated among individuals and organizations concerned. The programme officers solicit the opinions of colleagues and other stakeholders about the first draft. This is an important process, especially if the evaluation is directly connected to a decision that has a direct impact on vital interests of powerful stakeholders.

The draft report is primarily read by those directly concerned by the project. It is not given a wider circulation. Usually, it is photocopied in 15–20 copies and circulated among a limited group of people. They normally read and comment it on an individual basis. Usually, comments are mailed or faxed either to the programme officer in charge or directly to the evaluators. Sometimes internal meetings are arranged at the department. These normally involved only the programme officers and their colleagues, but sometimes also the implementation consultants. It is less common to arrange open seminars where all concerned stakeholders can discuss the reports together. The picture that emerges is that there is a limited discussion between stakeholders about the evaluation. Those who really make an imprint on the draft report are limited to a very small group of people.

For the purposes of this study, we identify 11 categories of stakeholder. Table 3.3 shows how involved each category was in shaping the reports.

Table 3 3. Who were asked to comment on the first draft of the evaluation report?

	Asked	Answered	Response ratio
Department evaluation coordinator	0	0	0
Division head	22	18	82
Colleagues in the department	13	12	92
UTV	10	3	30
Other Sida departments	13	7	54
Swedish embassy / DCO	11	9	82
Recipient officials	14	10	71
Local project staff	6	5	83
Target group	0	0	0
Implementation consultants	18	14	78
Other	2	2	100

Some are asked more frequently than others about their views. The most important reference is the head of the department. 73 per cent of the respondents sent the draft to the departmental head. The consultants responsible for the implementation of the project was the second most frequent reference (60 per cent). The programme officers' colleagues in the department are sometimes (in 43 per cent of all cases) asked to give their views. Colleagues in other departments at Sida were consulted with equal frequency.

The Swedish embassy in the receiving country, and UTV at Sida, are asked to comment in one third of the cases. Stakeholders in the recipient country are not routinely consulted, but in roughly half of the cases, the evaluation was sent to the institution in the recipient country responsible for the project. Judging by the responses, it is uncommon for the programme officers to solicit comments from the local project staff, but it is possible that these are covered by the category "implementation consultants". Not unexpectedly, the ultimate beneficiaries of the support are never asked to comment on the evaluation.

The programme officers solicit the opinions of colleagues and other stakeholders on the first draft. Some reply.[4] Even if stakeholders are not regularly asked to submit comments, they nevertheless may take the request seriously and submit their comments. The response ratio provides a simple measure of how actively they are involved in the evaluation. Colleagues in the department, including the division chief, almost always give their comments on the draft manuscript. So does the Swedish embassies in the collaborating country,

[4] Table 3.3 may give a false sense of accuracy. The programme officers did not always know if someone had submitted a comment, since in some cases the comments would have been sent directly to the evaluators. Furthermore, there is also a quality aspect to consider Sometimes comments may be very rudimentary, sometimes they bear witness of a thorough critical examination

department heads and local project staff. At the other end of the spectrum we find UTV, which rarely comments on draft evaluation reports.

An evaluation that addresses such questions as "shall we continue with this project?" or "do we need to change the orientation of the project in any way?" provoke different responses than questions like "what can we learn from this project in order to do better in the future?"

How did the programme officers themselves react to the first draft? For somebody as close to the process as most of them were, one would expect them to read the draft with a critical eye. This was also the case: almost all (83 per cent) expressed some sort of criticism. What were they critical about?[5]

Summarising responses to open questions may sometimes be difficult. In this case, however, a common pattern soon appeared. Generally, the critique focuses on formal aspects such as poor language, unclear outline and vaguely formulated conclusions and recommendations. Several respondents criticise the analysis for lack of depth. Many felt that it failed to provide any interesting and reliable results, particularly with regard to impact and cost-effectiveness.

This does not mean, however, that the officers were prepared to actually challenge the type of analysis made, or the facts and figures presented by the evaluator. In a few exceptional cases, the programme officers requested the evaluator to make substantial changes in the analysis and data material. But it was more common to record a dissatisfaction with the results while accepting the work that had been done. This pattern of behaviour can be explained in various ways. First, it may suggest a weakness in the typical evaluation procedure, particularly when more analytically demanding issues are being evaluated. When the draft report is presented it is usually too late to improve the analysis made by the evaluators. Second, it may be that the programme officer did not have a clear idea about the analytical requirements of the issues raised in the terms of reference. Hence, disappointment is likely to occur. The evaluators, on their part, may have lacked in professionalism. In order not to rock the boat, they refrained from challenging the core of the evaluation analysis.

Disseminating evaluation findings: the final report

A major purpose of an evaluation is to provide relevant and timely input to a decision-making process. This presupposes that the final evaluation report is made available to a wide range of interested parties. But does this commonly take place? How are evaluations disseminated and discussed by the various stakeholders and other interested parties? Table 3.4 summarizes the responses from our survey.

Table 3.4 Dissemination and discussion of the final evaluation report

	Yes	No	Do not know	Total responses
Report quality undisputed?	14	4	12	30
Report distributed to external parties?	14	13	3	30
Widely read and discussed?	18	8	4	30

[5] The respondents were asked an open question (number 3 7 in the interview protocol) "Did you have any critical comments on the draft?"

With regard to dissemination, it is clear that evaluations are not intended for a broader public. Draft reports in particular have limited circulation, as they are primarily meant for those involved in the project. But to what extent is the final report distributed to persons outside the core group of aid practitioners?

In about half of the cases the final report was given wider circulation. It was distributed, or at least made available, to other interested parties than the immediate stakeholders. But even if re-edited for a wider readership, evaluations are not known to have a mass appeal. It may be assumed that the main interest comes from an "extended family" of persons involved in development aid.

Did the programme officer feel that the quality of the final report was uncontested? The replies to this question are interesting. 14 respondents state that the quality was not disputed. Only 4 state that the evaluation was controversial and caused conflicts among the stakeholders. But as many as 12 respondents state that they do not know. This suggests that there is not much dialogue/discussion between stakeholders on the findings of an evaluation.

Do the programme officer feel that the report has been widely read and discussed? In most cases they do. However, since the respondents rarely qualify their statements, it is difficult to draw more general conclusions. When they do, the responses go in different directions. Some indicate that the circle of active readers has widened: "SAREC´s research council read and discussed the report and its recommendations". Others suggest that the external parties are not really interested: "The evaluation report was prepared for SAREC´s research council, but no one in the board read it".

Acting on the evaluation findings

The process behind an evaluation, from launching the idea and drafting the terms of reference to drafting and disseminating the final report, takes time. One would be led to believe that the evaluation is not very well synchronized with the requirements of the decision making process. Is there a risk that the information and recommendations contained in the report lose their usefulness, or can they be translated into concrete action

The evaluations provide, first and foremost, an analysis of the performance of the project. The analysis focuses on activities and outputs rather than objectives higher up in the intervention logic. This is regarded as important and useful in itself by the programme officers. But equally interesting is of course whether the evaluation has had any direct impact on the project. Did it result in any immediate and concrete actions? Table 3.5 summarizes the responses to a set of questions revolving around this issue.

Table 3.5 Were the recommendations accepted and applied?

	Yes	No	Do not know	Total responses
Were the recommendations accepted?	17	6	7	30
Did they result in concrete actions?	22	5	3	30
Is there a likelihood of any future actions?	12	11	7	30
Actions other than those recommended?	7	16	7	30

In about half of the cases the recommendations were clearly accepted. It is noteworthy that in six cases not all the stakeholders approved of the recommendations. (Where asked to elaborate on this, the respondents surprisingly often note that it is the implementation consultants that found it difficult to agree with the conclusions and the recommendations.) In as many as 7 cases the respondent do not know. This probably reflects the fact that the evaluation seldom was thoroughly discussed among the stakeholders. It was seen as a document for Sida to act upon, and for the programme officer it was less relevant what other stakeholders thought about it. In some cases the programme officers were new to the project and had for that reason not really discussed it with any of the other stakeholders.

It is a bit surprising to see that such a large proportion of the evaluations (73 per cent) do lead to some action, especially considering the long gestation period and the fact that not all of the recommendations (only 57 per cent) had been accepted. The apparent contradiction may be explained by the fact that those with critical views perhaps do not wield much influence on the issue of whether or not a recommendation shall be implemented.

Despite of the length of the whole evaluation process, normally 3–12 months, the recommendations were often of such a nature that they could be translated into concrete action. But what kind of action is taken? The material does not allow us to classify the actions into distinct categories. Instead, the following extracts may serve as illustrations:[6]

> The evaluation concluded that the collaboration between the Indian researchers did not work. In the future, SAREC will therefore give support to only one institution. The evaluation also recommended continued support along the same line

> The reference group has been given a more active role The programme officer is now trying to integrate the marine projects with the rest of Sida (the coastal initiative). Phase II will be implemented.

> One project will receive support for another three years The other project will be terminated within two years. The evaluation stresses the importance of accounting for results and recipient ownership, which are two recommendations taken on board by the officer

Sometimes it is not the recommendations that lead to action, but the report as a whole:

One sub-project was given a much smaller budget allocation due to the severe criticism in the report (although not a direct consequence of the recommendations). The evaluation is used to a large extent by the ministry when formulating the new agreement.

Two conclusions may be drawn from these samples. First, the recommendations are directed to Sida rather than to the recipient organization Second, they are not practical in the sense of addressing operational issues of relevance to the projects. They seem rather to raise issues of more strategic nature, in the medium- to long-term perspective This could explain why the recommendations, in spite of the lengthy evaluation process, still remain useful for the decision-making process.

The respondents were asked to assess whether the evaluation would result in any future action. One third of the respondents felt confident enough to state that the evaluation will continue to have an impact on future actions. Equally interesting was that just as many were quite certain that the evaluation will not result in any future actions.

---

[6] The extracts are picked at random from interview protocols. They are not necessarily representative of the whole population.

Unintended actions, that is, actions other than those recommended, were taken in a few instances. The following extract from one of the interviews serves by way of example:

> The evaluation had assessed a power plant which was part of the project, but not in use. This led to a direct intervention from the division head and the Director-General of Sida. They felt that it was unacceptable to support something that was not in use Rectifying this situation has now become a condition in the phasing-out agreement. This was not something that the report had recommended

4. The quality of the evaluation

Evaluations are undertaken in order to determine the merit or worth of a particular activity. The knowledge thus gained can be used to improve the performance of an organization in order to enable it to better reach its objectives. How the evaluation knowledge is put to use may vary: to increase the control of the organization, to increase learning within the organization, etc. A good evaluation is one that produces knowledge which is accurate, relevant and useful for those who are responsible for producing the output, as well as the intended beneficiaries. In order to fulfil these expectations, an evaluation has to meet certain quality standards: *utility, feasibility, propriety and accuracy.*[7]

Utility standards guide evaluations so that they will be informative, timely and influential. They require evaluators to acquaint themselves with their audiences, define the audiences clearly, ascertain their information needs, plan the evaluations to respond to these needs, and report the relevant information clearly and in a timely fashion.

Feasibility standards recognize that evaluations usually are conducted in a natural, as opposed to a laboratory, setting and consume valuable resources. Therefore evaluation design must be operable in field settings. Evaluations must not consume more resources, materials, personnel or time than necessary to address the evaluation questions.

Propriety standards reflect the fact that evaluations affect many people in a variety of ways. These standards are intended to facilitate protection of the rights of individuals affected by an evaluation. They promote sensitivity to and warn against unlawful, unethical and inept actions by those who conduct the evaluation.

Accuracy standards determine whether an evaluation has produced sound information. The evaluation must be comprehensive, that is, the evaluators should consider as many of the programme's identifiable features as is practical and should gather data on those particular features judged important for assessing the programme's worth or merit. Moreover, the information must be technically adequate, and the judgements rendered must be linked logically to the data.

These standards form a yardstick for us when we look at the quality of the evaluations produced by Sida.

Ideally, such an assessment should draw on two sources: the evaluation report itself and the evaluation process. The evaluation report provides the final proof of whether the standards were met or not; the organization of the evaluation process provides the conditions for the evaluation to meet these standards.

Our analysis of quality uses a comparative approach. On the one hand we have the results from the interviews with the programme officers. The same 30 evaluations have been independently assessed by one of the authors of this report (annex 5), and this forms the basis for the comparison with the programme officers' own assessments. The assessment format used for this database is found in annex 6. In other words, here we contrast the views of someone very close to the evaluation object with the views of someone who is neutral to it.

---

[7] American Evaluation Association, Joint Committee on Standards for Educational Evaluation, *The Program Evaluation Standards*, 2nd edition (Sage London [?], 1994), pp 5–6

It should be pointed out that the material has two major shortcomings. First, the questionnaire used in this survey was not directly designed for a full assessment of these standards. For example, we are not able to discuss all four quality standards in the same depth. One reason for this is that our survey is based on information collected within Sida. We have not been able to conduct the same survey among other stakeholders, most notably people and institutions in the recipient country. This reduces our ability to be specific about, in particular, utility and propriety standards. Second, the database from which we draw comparative data suffers from the same design limitations. It puts the evaluation report in focus, and has less to say about the nature of the evaluation process as such.

Quality as perceived by the programme officers

Once data collection is completed and the first analyses are presented, there are a series of stages involving comments and revisions of the draft report before a final text is ready. The programme officers' reactions/assessment of the report can therefore be expected to change over time. In order to capture this process we tried to focus on two points in time – when the first draft arrives and when the final report is presented. A broad picture of the programme officers own assessment of the final report is given in table 4.1.

Table 4 1 Satisfaction with the final evaluation report

Is the final report satisfactory in respect of:	Yes	No	Cannot tell
Language, clarity and organisation	27	3	0
Methods and data sources	26	4	0
Reliability	27	2	1
Objectivity	26	2	2
Practical usefulness	20	8	2
Wealth of information	17	13	0
New ideas	20	8	2

The picture of the evaluation process that emerges in table 4.1 is quite clear. In some contrast to the critical attitude expressed by most programme officers towards the draft report, a majority are clearly satisfied with the final evaluation. Obviously much happened between the draft review and the presentation of the final report.

In general, the programme officers felt that the evaluators were very receptive to the comments on the draft reports, and that their comments and suggestions were taken into account. By and large, almost all respondents felt that the final report was true to the intentions of the terms of reference. Disappointments were recorded, as noted above, mainly regarding analyses of impact and cost-effectiveness. In cases where the terms of reference explicitly require a discussion of impact and cost-effectiveness, the evaluators had encountered problems in presenting a convincing analysis. These problems were of such a nature that they could not be rectified in the period between the draft and the final reports.

Overall, the programme officers considered the reports well-written and well-organized. They find the information easily accessible. The evaluations almost always present the methods applied and data sources used in such a way that it is clear to the reader how the results have been obtained. The results are felt to be very reliable. The evaluations also gives a balanced and objective view of what has happened in the evaluated project. The interests and views of the various stakeholders are represented in a fair way.

However, the programme officers are less enthusiastic with the practical usefulness of the recommendations of the evaluation, although a majority feels that they are clear and implementable. They also feel that there is a logical link between the analysis of project/programme results and the recommendations made. However, when asked if the evaluation provided new knowledge about the evaluated activity, some 40 per cent responded negatively. The programme officers seemed to possess a fairly deep knowledge of the projects, to which the evaluations often had little to add. On the up side, when asked if the evaluations presented new ideas and a new perspective on the evaluated activity, about two thirds of the respondents acknowledged that this was the case. The evaluation also stimulated them to reflect on other issues in the area of development cooperation.

In order to shed more light on their assessment of the evaluation, the respondents were asked to elaborate on what they saw as the strengths and weaknesses of the evaluation. The replies vary considerably, and it is difficult to aggregate the answers into distinct categories. Instead, by way of illustration, table 4.2 pairs off strong and weak points for selected evaluations, as formulated by the respective programme officers.

Table 4.2. Weak and strong points of the evaluation

Strong points	Weak points
"A penetrating study, which gives a good general picture of the weak and strong points of the project"	"Analysis of impact is based on general discussion, rather than empirical based"
"The project has been well placed in a larger context"	"There are none"
"Concrete recommendations"	"It repeats itself"
"Fulfilled its function and played its role in the country strategy process"	"None really"
"Concise and to the point"	"A too general discussion about the projects' weak and strong aspects"
"Shows that quantitative improvements have taken place thanks to the project"	"Too little constructive critique"
"Provides a holistic picture of the projects within the road sector in the Baltic countries"	"It becomes a bit too extensive and repeats itself at times"
"Intra-disciplinary strong and gave useful recommendations not required by TOR"	Only concentrated on research results, not relating them to overall country context"
Careful analysis of the project from every possible angle"	"There are none"
"High reliability Some new ideas"	"Methodologically weak"
"Very positive, which makes it difficult to assess the real value of the evaluation"	"Weak language No scientific examination of the results. No discussion of country relevance of the project"
"Don't know"	"Never discusses in-depth results and effects"

There is no common pattern. What is regarded as a strength in one evaluation may be seen as a weakness in another. However, by way of crude generalization, most respondents identify the following characteristics as strengths: a project-specific orientation, a concentration of analytical efforts on project achievements, and a perspective that places the project in a country context. A weakness identified by several respondents is a failure to assess the impact of the project on the surrounding society.

Are the programme officers satisfied with the evaluations? This is not an easy question to answer. The picture that emerges is that of a process of approximation, whereby the evaluations by continuous improvements come closer and closer to the intentions of the terms of reference. While it is clear that the programme officers are acutely aware of the gaps and omissions in the evaluations, on the whole they seem content that the evaluators did their job and produced what the initiators of the evaluation regard as a useful product.

Using a database of aid evaluations to establish a yardstick for quality

What we mean by quality varies depending on who we are, where we work and, generally, what we believe our true interests to be. This is particularly true when it comes to such an elusive thing as an aid evaluation. The quality of an evaluation cannot be determined by simply referring to an "objective" yardstick. What somebody feels to be an accurate, feasible and useful evaluation can by somebody else be given a totally different interpretation. The programme officers in our sample did, generally, attach a great quality value on "their" evaluations. But what would happen if we looked at them from the perspective of the outsider, by someone who is not at all connected with the particular aid activity being evaluated?

The team members made a quality assessment of the same 30 evaluations, following the format developed and used by SASDA[8] and later by ourselves for the database on aid evaluations (see annex 6). In order to simplify the analysis, we have selected only those parts of the assessment format which have a direct bearing on our four quality criteria. This procedure raises a method question. is it possible to fully analyse utility only through the study of an evaluation report? The answer is a simple no. But within the framework of this particular study, this was the only feasible option.

Utility

A good evaluation should provide information that is relevant and useful to stakeholders in the project. Furthermore, the information should be timely with regard to the decision making process This presupposes that the evaluators understand the needs of their "audiences" - who they are, their interests and information needs. The starting point for at all being able to address the needs of the stakeholders is of course to talk to them during the course of the evaluation. Table 4.3 summarizes how well the evaluators have ascertained the needs of the most important stakeholders through interviews.[9] Our material does not allow us

[8] The Secretariat for Analysis of Swedish Development Assistance (SASDA) sorted under the Swedish Ministry of Foreign Affairs The unit was closed down in 1995

[9] In table 4 3 and following tables, "Exemplary" means that the evaluation has identified all members of this group of stakeholders and conducted in-depth interviews with each of them "Adequate" means that only the core members of each stakeholder group was identified and

to gauge the extent to which this has been done. However, the evaluation reports give information on who were contacted and interviewed by the evaluation team. Although still not saying anything about whether "needs have been ascertained", it shows whether a particular stakeholder has been given the chance to express his or her views.

Table 4 3 Who were interviewed by the evaluation team?

Interviews of	Exemplary	Adequately	Minimally	Not at all
Project staff	20%	57%	20%	3%
Beneficiaries[1]	27%	7%	37%	30%
Agency staff	20%	17%	23%	40%
Other donors	3%	10%	13%	73%

[1] The category "beneficiaries" includes various groups in the recipient country: government institutions, other local institutions, the target group, etc

In the table, stakeholders are categorized in four broad groups. project staff, beneficiaries, agency staff and other donors. It can be argued that this is too broad a classification. A more detailed breakdown would yield a more interesting analysis. This is true, but the rather uneven quality of our material – the evaluations – do not permit much else. The evaluations only occasionally contain information that would allow a more detailed stakeholder profile.

The table show that the evaluators primarily seek to ascertain the views of project staff. In about three quarters of the evaluations, project staff have been the primary interview subjects. Beneficiaries and agency staff members have received much less attention. This is perhaps not surprising. Evaluations are project-oriented, and it is therefore natural that evaluators concentrate on those stakeholders closest to the project. That the beneficiaries are not very important for the evaluators is not surprising. It confirms what we know about aid evaluations as a mainly donor-oriented activity. It is, however, more surprising to find that in 40 per cent of the cases agency staff is not among those interviewed. Perhaps it can be explained by the fact that the views of the agency staff can be communicated in other ways than through formal interviews. It is also possible that for reasons of objectivity, agency staff is not interviewed. Finally, few evaluator teams conduct interviews with other donors. Does this mean that they are not interested in the experiences of other donors? We can not give a firm answer to this question, only some possible interpretations. One, the evaluators may be uninterested because the agency commissioning the evaluation is not interested. Consequently the terms of reference do not ask for this kind of information. Two, they may be interested but there are other ways of ascertaining the experiences of other donors, for example through the use of documentation.

An indicator on utility can also be obtained by looking at the recommendations and for whom they are written. The pattern we see in table 4.4. confirms the information we have from the interviews. Evaluations are primarily produced for internal consumption and use within the donor agency. In less than half of the cases do we find recommendations suggesting actions to be undertaken by the recipient organization(s).

---

subjected to in-depth interviews "Minimal" means that only a handful of stakeholder representatives were interviewed, and not always in-depth

Table 4.4   To whom are the recommendations directed?

	Yes	No	Unknown
Recommendations for donor	90%	10%	
Recommendations for recipient[1]	43%	57%	

[1] The term recipient is used in a very broad sense, i.e including anybody from the recipient country . This could be a local consultant, someone from a government ministry or agency or even a representative of the target group.

Evaluations are useful for the donor and for those working in the project. They are probably less useful for the recipient organization and the beneficiaries of aid. However, this must remain an open question as we do not know anything about how the recipient consumes and uses the findings of an evaluation. This does not mean that they are not affected by an evaluation, nor uninterested. Nor does it mean that they do not have an opinion of the project in question. It is just that they appear to be marginalized in an evaluation process which is donor-driven. For example, it is noteworthy that recipients rarely initiate evaluations of aid projects.

To what extent does the 30 evaluations in our sample meet high utility standards? In a general sense they are probably useful since they are normally produced in time and in a language (English) that can be understood by most, if not all, stakeholders. More specifically, relying on our broad utility indicators, we suggest that it is the donor who stand to benefit most from the evaluation. Is this good or bad? Should an evaluation be equally useful to all stakeholders? The extent to which stakeholders find an evaluation useful depends mainly on the constellation of power within and around the project. Powerful stakeholders can be expected to exercise an influence that may make the evaluation more useful to them than to other, less powerful stakeholders. This may be the real situation, but is probably not the ideal one.

In an ideal situation, the evaluation would serve as a communication arena where all major stakeholders participate and exchange views and experiences. This does not necessarily mean that an evaluation should be equally useful to all concerned. But the evaluation exercise should be of such a nature that stakeholders feel that they have been heard and that their interests have been taken into account.

Accuracy

Can we trust the evaluation? Are the main questions clearly formulated? Is the information and analysis reliable, does it measure what it was supposed to measure and are methods technically adequate and applied in a correct way? In our assessment format we try to determine accuracy by gauging the extent to which the evaluations discuss issues of reliability, validity and methods, and testing whether the hypotheses and assumptions are clearly formulated.

Table 4 5   Indicators of accuracy

	Exemplary	Adequate	Minimal	Not at all
Reliability	3%	7%	23%	67%
Validity	0%	3%	13%	83%
Methods	0%	23%	63%	13%
Hypotheses and assumptions	0%	3%	27%	70%

The picture that emerges in table 4.5 is not very encouraging. It is rare for evaluators to use a self-critical approach and discuss the reliability and validity of their findings. It is more common to avoid these issues altogether. For a researcher it is mandatory to clearly state the working hypotheses and any assumptions made regarding the research task. For an evaluator this does not seem to be the case. The following statements give examples of how hypothesis and assumptions can be constructed:

> Our hypothesis is that all evaluations include adequate presentation of their methodology

and

> We assume that an adequate presentation of the methodology will lead to evaluation reports that have a high accuracy.

More common, although still not done in an exemplary way, is to have some sort of discussion on the methods used. One example suffices to illustrate how issues of methods are commonly treated in the evaluations reviewed here:

> The evaluation is based on documentation of the programme, scientific reports produced by the programme, and extensive discussions with persons concerned

An equally important accuracy issue is how the information used in the evaluation is presented. Again, in academic research it is mandatory to describe the information used and to give proper source references To ensure accuracy, we feel the same demands should be put on evaluations. Table 4.6. shows if questionnaires, examples of raw data, etc., are included in the evaluation report.

Table 4 6 More indicators of accuracy

	Yes	No
Questionnaires	17%	83%
Raw data	43%	57%
Interview protocols	13%	87%

In just under half of the evaluation cases the raw data is presented in the report. On the other hand, it is uncommon to include templates of questionnaires or interview protocols used by the evaluators. This creates difficulties for the reader, who is not able to judge the value of the reported opinions since they cannot know what kind of questions were asked and in what context Although evaluations generally do not qualify as academic research, there are still some basic research requirements that should apply also to evaluations.

Feasibility

A feasible evaluation is one that is appropriately designed for field conditions and makes the optimum use of available human and financial resources. There should be a balance between the cost of an evaluation and what it takes to answer the questions posed in the terms of reference. This is a quality standard that is difficult to measure accurately. There is no yardstick against which we can measure whether an evaluation delivers the goods at the most competitive price. Furthermore, our primary data sources also introduce limitations. Neither the questionnaire nor the database were primarily designed to facilitate an analysis of feasibility. We can therefore only offer a few broad indicators on feasibility the cost and timing of the evaluation, and the relationship between the cost of the evaluation and the cost of the project.

According to our respondents, an evaluation can take 3–12 months, sometimes more, from start to finish. The evaluation normally required a considerable input in terms of the programme officer's own time But what do we know about the total cost of the evaluation itself? Table 4.7 provides measures for the programme officers own perceptions in this regard.

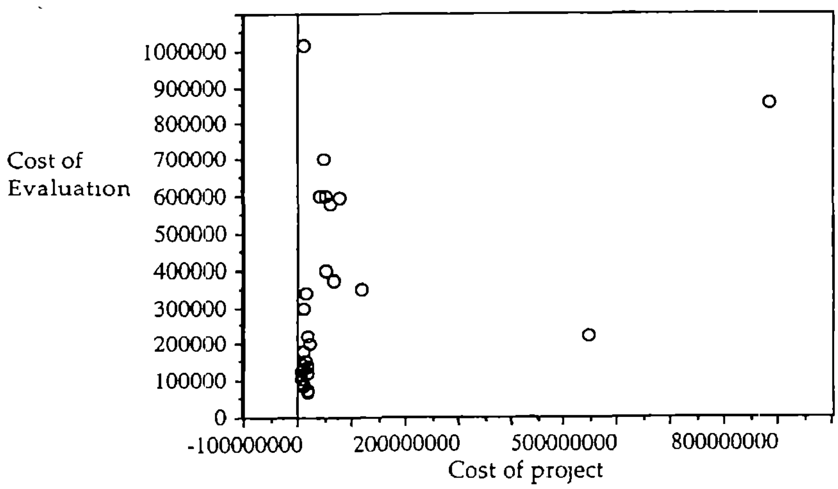
Table 4.7 The cost of an evaluation

	"Low"	"Medium"	"High"
Cost of evaluation	53%	31%	15%

Just about half of the evaluations are rated as having a "low" cost, which by the convention defined for the purpose of this study means that they required less than three man-months and one journey outside Europe. A "high" cost" means more than eight man-months and three journeys outside Europe. Evaluations in this category are not very common; the figure probably refers to larger evaluations of a thematic character. A "medium" cost means something between four to seven man-months and two journeys abroad. In other words, a good majority of the evaluations in our sample can be estimated to have a unit cost of SEK 200,000–500,000.

Can one expect a correlation between the cost of an evaluation and the size of the evaluated project in terms of cost? One could assume that the larger the project, the more expensive the evaluation. We tested this assumption in a simple scattergram in which we plotted Sida's cost for the evaluation against Sida's cost for the project.[10] Figure 4.1 shows the results.

Figure 4.1 Cost relationship



[10] This represents only the direct costs of the project and the project as they appear in the project budget and the evaluation budget For both the evaluation and the project there are indirect costs which do not appear in the budget For example, various costs incurred by recipient organisations and Sida's headquarter administration Neither is the time spent by people being interviewed, or people in recipient organisations trying to facilitate the work of the evaluators, assigned a monetary value.

Figure 4.1 clearly shows that the cost of an evaluation is not in any way related to the size of the project. It is probably the case that evaluation budgets are more influenced by factors such as the design of the evaluation in terms of method, the mix between foreign and local consultants, the cost of the consultants, etc., than by the cost of the project.

How are we to interpret these figures? Are the evaluations expensive? This is a difficult question that cannot answer with any acceptable degree of accuracy. Still, it seems reasonable to suggest that, in absolute terms, aid evaluations are not very expensive. One possible reference point is the project budget. Normally, aid evaluations account for only a fraction of the total project budget. Another yardstick could be the donor agency's total project and programme budget. We have no figures on this, but assume that evaluation activities account for a very small part of the operations budget.

Do the evaluations in our sample represent an efficient use of funds? This is different from asking whether they are expensive in an absolute sense. Cost-effectiveness links the cost for an activity to a clearly identifiable output. It is basically a comparative measure that establishes the cost-effectiveness of one activity by comparing it with another. To be able to analyse cost-effectiveness we therefore need two identical activities in terms of output. We argue that it is impossible to identify two evaluations that are sufficiently similar to allow a cost-effectiveness analysis. Applying such a concept to a research-related activity like evaluations therefore becomes meaningless.

Propriety

Propriety standards promote sensitivity to and warn against unlawful, unethical and inept actions by those who conduct the evaluation. Propriety is particularly relevant in aid projects. Aid is a relationship between two parties with unequal access to resources. This in itself introduces potential conflict between the parties. The very notion of aid – one helping another – introduces sensitive issues in terms of propriety. Furthermore, as evaluations are to such a large extent a donor-driven affair, one could suspect that the recipient party – rather than the donor agency – is felt to be under scrutiny. In order to balance the aid relationship, mutual respect is necessary. Hence propriety in evaluations becomes very important.

To what extent propriety standards are upheld in aid evaluation is difficult to tell. Our assessment format does not really address this issue. There are several indications in the material that propriety standards are at risk. First, because the initiative to start an evaluation is often taken without prior consultation with the recipient. Second, few evaluations are explicit in accounting for the data sources used and, in general, how they have arrived at their conclusions. Third, as table 4.8 shows, the evaluation teams generally do not include members from the recipient countries.

Table 4 8 Composition of the evaluation team

	Yes	No	Unknown
Donor present in evaluation team	0%	93%	7%
Recipient present in evaluation team[1]	20%	73%	7%
Sector expert present in evaluation team	80%	3%	17%

[1] The term recipient is used in a very broad sense, i e , including anyone from the recipient country. This could be a local consultant, someone from a government ministry or agency or even a representative of the target group.

The composition of the evaluation team is largely based on the idea that an evaluation should be made by an external, independent party. As a consequence, donors are not represented on the teams. These are rather dominated by persons – normally men – with specific technical competence. Recipient countries are represented on the teams in 20 per cent of the cases – however, not by persons representing a recipient counterpart organization or the target group, but by local consultants. It is an open question as to what extent their presence provides any guarantee that propriety standards are upheld. But – hopefully – at least they represent a possibility for other stakeholders in the recipient country to have their views and/or positions understood and respected.

Summary

It is striking how different the two quality assessments really are. The programme officers are of the opinion that after the draft report has been revised, the final product is satisfactory. To facilitate a comparison with our assessment format, we try to interpret the programme officers response according to our four quality criteria. We have pointed out above that our primary data sources suffer from weaknesses that prevents us from making a thorough comparison of all quality criteria. Three quality criteria – utility, feasibility and propriety – are particularly difficult to analyse. Our material is simply not good enough. This has to do with the type of questions asked to the programme officers, as well as the format used for our database. Equally important is the fact that we know very little about how stakeholders on the recipient side assess the various aspects of an evaluation.

We are on more solid ground when we assess *accuracy standards*. Here the response from the programme officers is pretty straightforward. In a majority of cases, the evaluation meets basic accuracy criteria. The report provides an impartial and broad picture of the project. It does not favour one stakeholder at the expense of others. All stakeholders have had an equal opportunity to present their views and opinions, and, furthermore, their views have been reflected in the report. The findings of the evaluation are considered reliable. Methods and data used have been clearly accounted for, and it is easy to understand how the evaluators have arrived at their conclusions.

Our assessment arrives at a different conclusion. The analysis suggests that the evaluations do not fulfil expected quality standards. This is particularly so when it comes to accuracy. An objection would be that our assessment is subjective, and that another assessment could arrive at completely different conclusions. Such differences may be expected when the task is to judge whether a given performance indicator rates as exemplary, adequately or minimally. Here a clear judgement may be difficult to make, and one could expect variations between different "assessors". However, the possibility of bias is reduced if the task is only to judge whether or not a given quality criteria has been met. It is apparent from table 4.5 that as far as reliability and validity is concerned, in a majority of cases they are not treated at all. Thus, we are fairly confident in concluding that the evaluations in our sample fall short of meeting the given accuracy standards.

An interesting aspect of quality is the gender composition of the evaluation teams. Most evaluators are men. Women are seldom recruited as evaluators, unless the project concerns issues that are traditionally regarded as of interest to women. child health, family planning, women in development, etc. This reflect an attitude towards gender within Sida that is difficult to reconcile with the importance Swedish aid attaches to gender equality in the recipient countries.

But would an increased proportion of women in the evaluation teams increase the quality of the evaluations? The question is difficult to answer without more thorough analysis of the evaluations in our database. However, an indication is given when comparing the variable "overall opinion" with the proportion of women in the teams. What we find is that (*a*) two of the three evaluations regarded as "excellent" were conducted by teams with women as team members; (*b*) of the thirteen evaluations that were regarded as "good", more than half included a woman in the evaluation team; and (*c*) none of the nine evaluations that were regarded as inadequate had a woman in the team. While these examples should be interpreted with much caution, they indicate that it may be worthwhile to investigate further the links between evaluation quality and gender.

A high *utility standard* requires an evaluation to be responsive to the needs of the "audience". This would seem to be the case for the evaluations discussed here, as the recommendations of the evaluations were, to the programme officers best knowledge, accepted by all stakeholders. Also, they were regarded as practically useful, resulting in concrete actions. They have not always provided new information about the project, but they have offered new perspectives. However, these affirmations do not provide a strong basis for any firm conclusions. A full understanding of utility would require that the views of more stakeholders were incorporated into the analysis. There are indications in the database, as well as from the questionnaires, that important stakeholders do not participate in the evaluation exercise. If lack of local participation can be confirmed by more extensive studies, it is clear that the evaluations fall short also in terms of the utility standard.

Regarding *feasibility* and *propriety standards*, we can only provide indirect hints as to how one could assess these two. We do not wish to draw any conclusions from the limited information found in the two data sources.

To conclude, the two data sets contain different assessments of evaluation standards. The most important, and clearly the most visible one, concerns accuracy. The responses to our questionnaires suggests a high accuracy. The database information indicates a low accuracy. How can we explain this difference? How can two independent assessments arrive at completely different conclusions? Why do the programme officers rank the evaluations so highly in terms of accuracy, while our more detailed assessment suggest that they fall short of meeting basic quality indicators? One explanation could be that the officers are such an integrated part of the whole evaluation process that they tend to identify themselves with the final product. They launched the initiative, participated in selecting the evaluators and monitored the evaluation process from beginning to end. Sometimes they are also quite close to the project being evaluated. They know it very well, they know what has happened, etc. It is quite plausible that this "closeness" gives them a perspective on what is "reliable" and "valid" information that differs from that of an external analyst.

5. Does practice differ from policy?

Let us finally consider how these findings relate to Sida's evaluation policy. Does the policy have any practical impact on the conduct of evaluations? Does the actual practice follow the ideas and intents expressed in the policy, or is there a gap between policy and practice?[11]

The interviews with programme officers and evaluation coordinators show that all were aware of the existence of the policy. However, not many had consulted the policy when writing the terms of reference for their evaluations. A reason often given was that the policy did not exist at the time. But, as far as we know, the policy has been in existence since late 1995, and several of these evaluations were initiated in late 1995 or early 1996. Consequently, the statement cannot be true in all cases. The answers provided during interviews more likely reflect that the respondent recognized that the policy should have been consulted, but that it had been forgotten, or the connection was not made at the time.

The responses to the interviews suggest that the establishment of the policy has not had any major impact on the conduct of evaluations. None of the respondents suggest that the policy has led them to do other things than they otherwise would have done, nor does anybody claim that they have developed or changed their "normal" way of conducting the evaluation process.

To what extent, then, does practice differ from policy? The implementation of the policy – or rather the lack of it – is fully visible throughout this report. Each chapter and each section can be compared to what the evaluation policy has to say about the conduct of evaluations. In many ways the policy codifies what most people would agree to be desirable practice, but, for different reasons, reality is different from the ideal. The major differences between policy and practice seem to emerge in the following areas:

1 In the preamble to the policy, there is a definition of what an evaluation is. Do the evaluations at hand comply with this definition? It appears to be impossible to arrive at any other conclusions. The definition call for "a systematic and thorough investigation". Who can say that these evaluations were not systematic, given the constraints that the evaluators faced in terms of budgets and methodological directives? But with a more absolute definition, we might have concluded that several of these reports simply do not meet the minimum standards for evaluation reports.

2 The policy states what the contents of an evaluation should be. "[T]he following aspects of the activity to be evaluated shall always be taken into consideration: . . . relevance, goal achievement, effects, efficiency, sustainability". The list can be compared to the findings of our quality assessment of the evaluation reports in chapter 4. There is clearly a gap.

3 Consider the purpose. The policy says that evaluations are started for "purposes of control, learning, and development of knowledge". But the reason why these evaluations were initiated rests on much looser ground. They were initiated because it was felt appropriate, it was timely or something was needed for the

[11] Note that this report only studies that part of the evaluation system which falls under the duties and responsibilities of operational departments. While it does not cover all of Sida's evaluations, it covers a major part, and thus the question of how it relates to policy is important

evaluation plan. Later this was developed into a more rational and explicit purpose, mainly in terms of managerial use. Compared to the policy, there is thus less of a focus on control, learning and knowledge development as reasons to evaluate.

4 The policy explicitly says that evaluations should be undertaken with due "concern for objectivity and impartiality". There is no doubt that the evaluators were not directly involved in the implementation of the projects they were evaluating. But they seem to have had long-term relationships to Sida. Most were well-known to the programme officers and had in several cases been engaged in similar tasks many times before. The question is whether a closer examination of their relation to the aid programme would confirm their objectivity and impartiality.

5 The policy specifies that the "partners in recipient countries should be engaged in the evaluation process; the evaluations should be undertaken in a spirit of cooperation". The interviews revealed that the partners were involved at the inception of the evaluation in somewhat less than 40 per cent of the cases, that they were engaged in the process only seldom, and that they received a draft report to comment on in only 50 per cent of the cases. There are good reasons to assume that these figures should be less than 100 per cent, due to the nature of global projects and sometimes due to the issues raised. But there can be no doubt that the policy stipulates another and more ambitious level of cooperation with the partners.

6 Perhaps the most important difference between the policy and actual practice lies in the quality of the reports. The analysis above shows that the quality is low, and many times there are no good reasons to believe the evaluators. They do not present any raw data and they do not inform the reader how they have arrived at their conclusions. The methodological awareness is low; whether we believe them or not is a matter of how well we know them, and of how well their conclusions coincide with what we believe ourselves to know anyway. Evaluations are said to be subject to the same quality requirements that apply to all Sida activities. We cannot be sure what that means, as there are no quality indicators specified in the policy. The evaluations certainly do not reach very high on the quality indicators we have used, but on the other hand these have never been recognized or applied by Sida. However, there seems to be a need to develop quality indicators that can be used to assess the evaluations according to the intentions of the policy.

7 Finally, the policy also notes that "evaluation activities should conform to the requirement of efficiency". Decisions concerning an evaluation shall as a rule not be taken without an assessment of its value to Sida and the recipient in relation to its costs. The information from the interviews give us only vague indications here, but nowhere does it appear that any such considerations were made. On the other hand, nowhere is it explicitly stated that such consideration were never made.

6. Summary and conclusions

One of the consequences of the reorganization of the Swedish aid organizations in 1995 was the creation of a new evaluation function. This function is now called the Department for Evaluation and Internal Audit (UTV). Possibly the most interesting feature of the evaluation system at the new Sida lies in the combination of a widespread, even diffused, operational responsibility at the level of departments, and a centralized, semi-independent evaluation function.

The tasks of UTV fall in three categories:

1 to initiate and carry out evaluations under its own auspices or in cooperation with other organizations;

2 to review, examine for quality and improve Sida's evaluation system generally and to cooperate with the operative departments (for example by disseminating evaluation findings); and

3 to participate in international evaluation work to exchange experience and cooperate with the evaluation functions of agencies, and to support the development of evaluation capacity in recipient partner countries.

One of the thematic areas of work of the UTV concerns organizational change and learning. The study presented here forms part of this theme. In short, the purpose is to map the use of the evaluation instrument in Sida, by answering the following questions:

1 How and why are evaluations initiated?

2 How is the evaluation process managed, from the setting of a purpose, the decision to evaluate and the commissioning of a study?

3 How are the results of this process used?

4 Do the evaluations meet acceptable standards of quality?

The study answers these questions in relation to evaluations conducted by the operational departments. It does not analyse the evaluations undertaken by UTV itself.

The study builds on interviews with 9 evaluation coordinators, as well as 30 programme officers who have recently been in charge of an evaluation process. The quality analysis covers more than half of the evaluations completed during the past fiscal year, with a focus on those 30 where it was also possible to interview the programme officers in charge. The data was gathered in three separate processes, described in annexes to the main text of this report.

How and why are evaluations initiated?

Sida is a very decentralized organization and consequently it is usually the programme officers who initiate an evaluation. It is rare that an external agent, or any other part of the organization, pushes for an evaluation. The formal decision to evaluate is taken by the head of the division, and thus he or she also is involved at an early stage.

The most common reason to initiate an evaluation is that it had been decided in the project document, or it was part of the agreement with other stakeholders, that an evaluation was to take place after some years. In some areas it is simply common practice to evaluate a project after a given period of time. Some evaluations were initiated because they seemed to be interesting to have in the evaluation plan, and some because of a need for input to country strategies. Yet another reason to evaluate was that the project had come to an end, or was approaching a mid-term review. In both cases, Sida would soon be entering negotiations on whether and how the programme should continue. Here an evaluation providing a review of results was expected to contribute to the decision-making process.

The initiative to evaluate appears to rest on loose grounds. There is nothing wrong in this *per se*, it lies in the nature of an initiative to be vague. But how is the initiative processed into a more carefully considered purpose? In the process that follows other stakeholders are brought into the picture, but the range of consultations appears to be very narrow. Most have informal contacts with their colleagues in the division. But only about one third of the respondents consult the recipient partner country at this stage, and about as many consult the embassies or the project consultants. Only one had contacted the department's evaluation coordinator, and five had contacted UTV.

The immediate and direct purpose of an evaluation is defined in the terms of reference, and this document often appears to be largely copied from the evaluation manual. Most consult the manual during the process and when preparing project documents and reports. But few look for possibly challenging sources such as other evaluations, other project experiences or other donor reports. We found a lack of practical focus at the level of the individual project or programme. Instead of a formulation of purpose the terms of reference often include what may best be described as a shopping list. The link between a vague initiative to a practical and concrete purpose to guide the evaluation thus appears to be weak.

How is the evaluation process managed?

The role of the programme officer in charge of an evaluation appears to be a lonely task. He or she has few contacts with others in the organization, and those they have are generally the fruit of their own informal and personal networks. But as most have long working experience in the organization, one may assume that such networks are well-developed. Only in 2 of the 30 cases had the respondent formed a reference group for the evaluation.

All evaluations were commissioned to external consultants. These were largely found with the assistance of colleagues, and most were well-known to Sida from earlier evaluations. That is, they form part of a group that is commonly used for tasks such as these.

Only one programme officer claims to have used a consultancy roster at UTV (although UTV does not keep such as a roster). Nevertheless most programme officers state that it was easy to find qualified evaluators. Only rarely were evaluators selected on the basis of competitive bidding – in fact this only happened on 5 occasions (17%). This is more surprising as it seems that several evaluations cost more than SEK 200,000, a cost ceiling above which there is a requirement to place the assignments on tender.

The evaluations are usually completed rapidly once they have been commissioned, often within a few months. But the process from the first initiative to evaluate and up to the dissemination of a final report is a more lengthy process. Nevertheless all the evaluations here were completed within the planned period. It seems as if the evaluations arrive in time to be useful, though we cannot be quite certain of that. The programme officers spend a considerable amount of time on the evaluations, often between three and four weeks full time work.

How are the evaluation results used?

When the draft report is received, it is circulated – generally speaking to the same group of stakeholders that were consulted in connection with the initiative to evaluate, that is, a fairly small group of people. In almost 50 per cent of the cases a copy of the draft was sent to the recipient country authorities for comments. The consultants who were engaged in the implementation of the project in most cases also received the report for comment.

Comments usually focused on editorial issues, but also often gave expression to some disappointment regarding the depth of analysis. The recommendations were often found to be vague and difficult to act on. But the reports were mostly changed in response to comments on the draft report. As a result, when presented with the final report, programme officers generally felt that the product had been improved considerably.

Of the 30 evaluations examined here, the programme officers confirmed that the recommendations had been accepted in 57 per cent of the cases, confirmed that the recommendations led to concrete actions in 73 per cent of the cases, and in 40 per cent of the cases assumed that they would be likely to lead to actions in the future. In addition, in 23 per cent of the cases the evaluations led to actions other than those recommended.

Are the evaluation reports good?

The quality of evaluations is an illusive subject, and it must be recognized that what one person regards as a good and reliable report can by another person be given a totally different interpretation. When assessing the quality of evaluations we have focused on the reports, but the report itself is only a manifestation of the whole process – which may contain quite different qualities. The approach here builds on two contrasting quality assessments.

First the programme officers were asked to judge the quality of the final report, following amendments to the draft. Most were pleased with the outcome of the evaluation process They thought the evaluators had arrived at reliable conclusions, and that the project was given a fair and adequate treatment. They thought the reports were clear, concise and well written. The reports were found to be relatively weak on new ideas and practical usefulness, but, nevertheless, two thirds of the respondents were satisfied with the evaluations in these respects too.

Second, the evaluation reports were assessed using a standardized worksheet containing 50 variables that address various aspects of quality. This was applied both to the 48 evaluations completed during the year and to the 30 evaluations that were followed up with interviews. It is the latter sample that we refer to here. The overall opinion was that 3 per cent were rated "excellent", 27 per cent "good", 53 per cent "adequate" and 17 per cent "inadequate". In particular, the reports were found to be methodologically very weak. Their analysis of crosscutting issues and sustainability left much to be desired. Cost-effectiveness is seldom assessed, less than 20 per cent analyse the aid agency's performance, and only slightly more than 50 per cent include an adequate analysis of the achievement of project objectives. An external, standardized assessment thus provides a very different picture from that of the programme officers.

References

Bruce, C. and Finzı, U. (1994), *Improving Monitoring and Evaluatıon in Swedish Development Assistance*, SASDA: Stockholm.

Burke Johnson, R. (1995), "Estimating an evaluation utilizatıon model using conjoint measurement and analysis", *Evaluation Review*, vol. 19, no. 3, pp. 313–38.

Carlsson, J., Köhlin, G. and Ekblom, A. (1994), *The Political Economy of Evaluation: International Aid Agencies and the Effectiveness of Aid*, Macmillan: London.

Carlsson, J. and Jörnmark, J. (1994), *Evaluation and Learning in BITS*, SASDA· Stockholm

Carlsson, J. and Forss, K. (1997), "The use of evaluation results in development aid". *Andante Occasional Research Paper Series,* vol. 97, no. 2.

Forss, K. (1994), *Planning and Evaluation ın Aid Organizations*, EFI/IIB· Stockholm

Forss, K. (1994a), *Några aspekter på kvalitén hos utvarderingar av svenskt biståndssamarbete*, SASDA: Stockholm.

Forss, K. (1994), *Evaluations ın Multılateral Development Cooperatıon*, SASDA: Stockholm.

Forss, K. and Carlsson, J. (1997), "The Quest for Quality - or can evaluatıon findings be trusted?" *Andante Occasional Research Paper Series*, vol. 97, no. 1.

Forss, K. and Laursen, J. (1997), *The findings and the quality of evaluations in the Nordic Development Fund.* A report commıssioned by NDF. Helsınkı.

Groȷer, A., Carlsson, J. and Forss, K. (1997), "Promotıng democracy: what do we know and how can we measure results?", *Andante Occasional Research Paper Series,* vol. 97, no. 3.

Joınt Commıttee on Standards for Educational Evaluatıon (1994), *The Program Evaluation Standards*, Sage: London.

Samset, K., Forss, K. and Hauglın, O. (1993), *Learning from Experience: A Study of the Feedback from the Evaluation System ın the Norwegian Aid Administration*, Norwegıan Mınistry of Foreıgn Affaırs: Oslo.

Stokke, O. (1994), *A Perspectıve on SAREC's Evaluation System: Polıcy and Performance*, SASDA: Stockholm.

USAID (1989), *Review and Qualıty of A I D Evaluatıons FY 1987 and FY1988*, USAID· Washıngton, DC.

Watson, C. (1991), *A Study of the Qualıty of CIDA's Bilateral Evaluations 1988–89 and 1989–90*, CIDA Ottawa

Annex 1

Terms of reference, Project description
Karin Metell

Table of contents

1. Terms of Reference

Terms of Reference for consultant assignment within the theme area "Change and learning at Sida: Using the evaluation tool"

Background

A survey of the use of the evaluation tool is being conducted within the framework of the theme area "Change and learning at Sida". The study draws mainly on interviews with evaluation coordinators and programme officers for evaluations commissioned by Sida's sector and regional departments in 1996, and on a review of evaluation reports with regard to certain quality aspects. Below are listed some of tasks to be performed by consultants as part of the survey.

The assignment

The assignment involves the following activities:

Activity:	**JC**	**KF**
Revising interview guidelines	1	1
Meeting with three interviewers (KM, LS, TS)		1
Conducting pilot interviews (1 per interviewer). KF participates in each interview.		1
Exchanging experiences after conclusion of pilot interviews		1
Revising interview guidelines		1
Exchanging experiences after conclusion of interviews		1
Writing final report (incl analysis of data and presentation)	15	15
Total number of days	16	21

Consultants, organization and time plan

The work is to be done by Kim Forss (KF) and Jerker Carlsson (JC) at Andante Consultants. Commissioning authority is the Department for Evaluation and Internal Audit. Project leader is Karin Metell (KM). Conducting interviews with evaluation coordinators and programme officers are members of the Department staff and contracted project staffers Lisa Segnestam (LS) and Tove Strömberg (TS). The work is conducted within the framework – and in accordance with enclosed description – of the project "Change and learning at Sida: Using the evaluation tool".

Time plan:	
Revising interview guidelines	– 14 Nov.
Meeting with three interviewers (KM, LS, TS)	15 Nov
Conducting pilot interviews	– 22 Nov.
Exchanging experiences after conclusion of pilot interviews	25 Nov
Revising interview guidelines	– 27 Nov
Conducting interviews	– 20 Dec.
Drafting report (incl analysis of data)	– 20 Jan
Final report	**31 Jan.**

Reporting

The final report shall present the results of interviews with the 10 evaluation coordinators, conducted and presented by UTV. The results of interviews with the approx. 40 programme officers managing evaluations, conducted in collaboration between UTV, LS, TS and the consultants, shall be collated and presented by the consultants in the final report. A quality review of evaluation reports commissioned in 1996 (approx. 40) is to be performed by each interviewer in preparation for the relevant interview. Data analysis of all quality reviews is to be conducted by TS. The results of this analysis, and a comparison with similar reviews of evaluations from previous years, are to be presented in a separate section of the final report.

A first draft of the final report in English is to be delivered to UTV no later than 20 January. After discussions with and having received comments from UTV, the final report shall be presented, in two camera-ready copies and on diskette, no later than 31 January. The consultants are also to provide, on diskette, any and all background data used in the study.

Project description

"Using the evaluation tool at Sida"

Background

Sida commissions a great number of evaluations each year, most of them in connection with the on-going follow-up of Sida development cooperation programmes and projects. Sida's Department for Evaluation and Internal Audit (UTV) is charged with monitoring the quality of evaluation activities. The proposed study is in furtherance of this work. It raises fundamental issues regarding the use of the evaluation tool in Swedish development cooperation efforts. The study forms a project within the framework of UTV's theme area "Change and learning within Sida".

Purpose

The purpose of the study is to chart how the evaluation tool is used within Sida. The study is to assess these efforts in light of Sida's overall evaluation policy. The results are to be of use in regular follow-up surveys of the use of the evaluation tool in the future. The study is also to suggest possible improvements regarding evaluation planning and implementation, and the feed-back of evaluation results.

Plan of activities

The survey covers evaluations scheduled in Sida's evaluation plan for 1996. Other evaluations conducted during the year may also be included. The survey will thus involve planned evaluations, both completed and non-completed, as well as evaluations initiated subsequent to the original plan. Several, partly overlapping, phases of the study may be identified at this stage:

1 Survey of existing literature, with special emphasis on similar studies by other donor organizations. The purpose here is to select and specify relevant areas of inquiry for the subsequent study.

2 Preparations. Formulation of queries. Drafting of interview form. Collation of data regarding the various departments' evaluation activities in 1995/96.

3 Interviews with Sida's evaluation coordinators The purpose of the interviews is to provide insights into departmental structures, policies and practices for planning, implementing, following up on and using evaluations.

4 Interviews with programme officers that in fiscal year 1995/96 had primary responsibility for the administration of one or more of the evaluations covered in the survey. The interviews will highlight issues regarding purpose, planning, consultant procurement, implementation and use.

5 Review of evaluation reports with regard to certain quality aspects. Information from Sida's data base is supplemented with quality criteria for evaluation reports. The reports are reviewed with an eye on content, issues raised, presentation and implementation method.

6 Results and recommendations to UTV and other Sida departments should be summarized in the first draft of the report.

7 Presentation and follow-up of conclusions and recommendations.

Organization

Primary customer of the project is the *Sida Board*, being the responsible authority for Sida's evaluation work. The study should also be of interest to departments within Sida that use evaluations as part of their work to monitor results and follow up on development cooperation programmes and projects.

The departments' *evaluation coordinators* will play an important role both as respondents to and customers of the study. Data will be collected by means of structured interviews with each coordinator. This may be supplemented by information regarding departmental structures and policies for evaluation work gathered through discussions with *unit* and *department heads*.

In additions, some *40 programme officers* managing evaluations conducted for Sida in 1995/96 will be interviewed. To allow programme officers to prepare answers to some of the questions, relevant parts of the interview guidelines will be distributed in advance.

Interviews with evaluation coordinators will be conducted by UTV officers. Interviews with programme officers will be conducted by UTV staff, assisted by contracted expertise. Results of interviews and of quality reviews of reports will be collated and analysed subsequent to each interview. The quantitative collation and analysis of data and the final reporting on the various components of the study will be done by outside consultants.

Limitations

The scope of the study is limited to the use of the evaluation tool at Sida. The role of the recipients is of course important in evaluation planning and implementation, and – not the least – in the use of evaluation results. However, such aspects are considered only to the extent that they are reflected in information provided in interviews with Sida programme officers. In other words, the focus is on the view and practical use of the evaluation tool on the part of Sida officers and managers.

Preliminary time plan

Activity:		Completion date
1)	Literature survey	completed
2)	Drafting of interview form, etc	– 15 Nov.
3)	Interviews with Sida's evaluation coordinators	–30 Nov
4)	Interviews with programme officers	– 20 Dec
5)	Quality review of evaluations conducted in 1995/96	
	a) data collection and preparation of data base	13 Dec.
	b) data processing and analysis	– 20 Dec
6)	Collation of conclusions and recommendations	– 31 Dec
7)	Presentation	– 15 Mar.

Personnel requirements

The project is led by UTV staff. Total effective time required by UTV staff is estimated to be 12 person weeks. Total time requirements for outside consultants is estimated to be 10 person weeks.

Annex 2

Interviews with Sida's evaluation coordinators
Compiled by Karin Metell

Table of contents

1. Introduction

In 1996, Sida's Department for Evaluations and Internal Audit (UTV) initiated an in-depth follow-up study of all evaluations ordered by Sida during the year. Falling within the theme area "Change and learning at Sida", the aim of the study is to establish how the evaluation tool is used within Sida. How and why are evaluations initiated? What is the process in evaluation planning, authorization and in the procurement of evaluation consultants? How does Sida make use of the evaluation results?

New Sida amalgamates what previously was five separate aid agencies, each with its own distinct administrative routines and practices. This applies not the least to evaluations and audits. To get an initial, overall perspective on the evaluation practices of the different departments, interviews were conducted with Sida's evaluation coordinators. This paper summarizes the results of these interviews.

All of Sida's 14 departments regularly conduct evaluations as part of the standard follow-up of on-going work. This study concerns only *evaluations of Sida-funded development cooperation*. It does not cover evaluations of a more administrative character, such as those conducted by the administration and information departments. Also excluded are evaluations commissioned by UTV, SEKA and POLICY.[1] This limits the number of departments covered in the study to nine. The interviews were conducted to gain an initial perspective on how the departments' evaluation activities are organized and on the rules or norms that guide them.

As the function of the evaluation coordinator was only recently instituted, some of the interviewed officials have limited experience of any but their own unit's evaluation work. While neither complete nor comprehensive, the interviews provide descriptive profiles of the various departments' evaluation work Information gleaned in the interviews is developed further in the subsequent survey, in which some thirty-odd programme officers engaged in evaluation work in the course of the year are interviewed. This follow-up helps shed light on interesting differences between departments' rules and practices.

The interviews followed a structured format (see Annex 4 to the main report). The interviews were conducted by the respective contact person at UTV for the different departments, together with the project leader for this study. Each coordinator was informed of the purpose of the survey but was not given the questions in advance or in any other way prepared for the interview.

A further aim of the interviews was to follow up on the role of the evaluation coordinator as such. The interviews therefore conclude with a set of questions on the tasks that the coordinators have performed so far, and on how they see their own role. The results of this follow-up is presented in section 5.

---

[1] SEKA and POLICY have not conducted any evaluations during the year

2. Sida's organization for evaluation and the role of the coordinator

Each department has appointed an evaluation coordinator. These individuals, together with UTV and a group of Sida advisers, constitute a network on evaluation issues at Sida. In broad terms, the role of the coordinators is, within the network, to represent their respective departments and, within the departments, to inform on contacts with UTV and the network. In practical terms, this entails the following responsibilities:[2]

- Keeping tab on departmental needs for competence-building in the evaluation area, and participating in the formulation of training programmes;
- Drafting and compiling annual evaluation plans for the respective departments, in keeping with established guidelines;
- Selecting evaluation efforts to be managed in collaboration with UTV, when preparing the annual evaluation plan;
- Accounting for the implementation of annual evaluation plans and other evaluation activities, in keeping with established guidelines;
- Assisting and advising the departments in evaluation issues, together with UTV.

[2] The following points are tasks that have been discussed at network meetings but never formalized. (See minutes of meeting dated 96–12–18)

3. Basic facts regarding the evaluation coordinators

The evaluation coordinators number fourteen to date. All were appointed by their respective departments in 1996. The position as evaluation coordinator was created at the end of 1995 in conjunction with the formulation of the 1996 evaluation plan. Only 7 of the original appointees remain as coordinators. In other words, several have been replaced and others have joined.

Of the 14 coordinators, 8 are women. Of those interviewed, about half have worked with aid issues for 20 years or more, either within Sida (SIDA) or within some other Swedish organization. Two persons have 10–12 years of experience with development work, and three persons have 5 years experience or less. All but one have at least one year's experience of field work. Within the past 18 months, six persons have been given new work assignments.

Most of the coordinators have extensive experience in administrating evaluations. This means that they have initiated evaluations, been responsible for the formulation of the terms of reference, selected and commissioned consultants or researchers for the task, and followed up on their recommendations and suggestions. Six out of nine have administrated evaluations for Sida or some other organization four times or more. The remaining coordinators have administrated evaluations 1–3 times in the past.

However, few of the coordinators have first-hand experience of the evaluation process as such, either as observers or as members of evaluation teams. Three persons state that they have participated in evaluations as observers or team members. Only one person has experience of participating in evaluation work in the field.

4. Departmental structures, policies and practices for evaluation

The use of the evaluation tool is assumed to vary between and within Sida's departments. In the regional departments RESA, REWA and REMA, implementation authority is for most projects delegated to either the sector departments or the embassies concerned. These departments rarely originate evaluations of their own. RELA maintains implementation authority to a greater degree, and regularly orders evaluations of the projects and programmes managed within the department. Sida-Öst and SAREC initiate and manage many evaluations each year. Finally, the three main sector departments, DESO, INEC and NATUR, all maintain implementation authority and together account for the bulk of Sida's evaluations.

Guidelines

There are no evaluation guidelines specific to given departments. Nor does it appear that sector-specific guidelines contain guidelines regarding evaluation activities. The routines and practices used have rather evolved from those of the previously independent aid agencies, none of which had any written policies on evaluation work.

Selection

To the question regarding what is/what is not evaluated, all coordinators reply that no formal rules exist. A common selection criterion seems to be that large, costly and lengthy projects and programmes are given priority over small ones. Programmes up for renewal are often – but not always – evaluated in anticipation of negotiations with recipient countries. This last criterion is in several departments problem-oriented, that is, projects where follow-up and reporting is satisfactory, and where project results are in relatively good order, are evaluated less frequently.

Initiative and authorization

The initiative to undertake an evaluation usually comes from the sector department (including Sida-Öst and SAREC) or embassy programme officer responsible for the project or programme in question. According to the respondents, it is uncommon that the initiative comes from unit management, which instead wields its influence through discussions at unit-level reviews of current projects and programmes. Often an evaluation is planned in agreement with the recipient country, and is initiated subsequent to annual reviews or other activities in the recipient country.

The fact that most initiatives are taken by project programme officers may reflect the fact that evaluations are often called for as bases for decisions regarding major changes in projects and programmes. Sida's regional departments have so far wielded very little influence in this regard, despite the fact that evaluations constitute one of the primary bases for their work in evolving country strategies.

The authorization to evaluate a project or programme is often made at a relatively late stage, when a consultant is contracted to conduct the evaluation. Authorization is often given by unit management.

Cooperation in the evaluation process

Cooperation in the evaluation process occurs mainly within the unit concerned. In most cases, the responsible department or embassy programme officer drafts the preliminary terms of reference. The influence wielded by unit chiefs seems to depend on the individual, but may be significant during the course of the work. The terms are not actually formalized until the evaluation itself is authorized, and they in practical terms already have been defined. Cooperation between units does not seem to be common, except in cases where units have recently been reorganized, and previous or current programme officers have been reassigned to other units (e.g. NATUR). This means that advisers above the departmental level, as well as the POLICY department, play a marginal role in the formulation of the terms of reference. Likewise, UTV wields little or no influence on the formulation process, the selection of consultant, etc. Certain formal cooperation takes place between sector and regional departments, but not on regular basis.

A common practice in many units seems to be to allow parties concerned in recipient countries to comment on the terms of reference of planned evaluations. This applies in particular to divisions from "old" SIDA and SAREC. The opportunity of recipients to influence the formulation of the terms of reference is seen as slight by all coordinators, except in the case of SAREC.

In short, evaluation preparation is handled internally as part of regular project work..Programme officers work individually, and personal contacts within or outside the unit determines who he/she cooperates with. There are no formal requirements regarding cooperation, and available advisory resources at Sida (within units or at POLICY and UTV) are only utilized to a marginal extent. Recipients have little influence on the process.

Consultant procurement

Who is consulted by Sida for evaluation assignments? There are no formal rules, but a principle applied by all departments is that the consultant should have no ties to the project concerned. His or her impartiality must not be in doubt. This is stressed by all coordinators.

Sida personnel may be used, though not to conduct evaluations of projects or programmes where they themselves are involved as programme officers. Some respondents were sceptical to the use of Sida personnel – and to consultants from recipient countries – as evaluation team leaders. In practice, however, it is very rare that Sida personnel participates in evaluation teams.

While some units with past links to BITS as policy only hires Swedish consultants, SAREC prefers to hire consultants from developing countries. SAREC's policy is motivated by the ambition to raise the general competence and capacity of these researchers and of the Third World research community as a whole At other units, where evaluations are conducted by teams of several persons, there is also the expressed wish to include consultants from the countries that Sida cooperate with. The main reason here is to exploit the regional and cultural competence that these consultants offer.

The evaluation consultant procurement process differs widely between departments. Within INEC and Sida-Öst, many evaluations are contracted out using competition bidding, for commissions below Sida's stipulated ceiling. Other departments, such as SAREC and the regional departments, rarely if ever use competition bidding as an element in the procurement of evaluation consultants.

None of the departments keep registers on available consultants. A few units keep their own resource base registers, but none list search keys that allow searches by evaluation competence. The use of consultant registers or similar tools to identify possible consultants seems to be uncommon. In addition to sector-specific competence of relevance to the given assignment, a common criterion for selection among the departments is past experience of evaluation work commissioned by Sida. References and recommendations from colleagues is therefore much valued.

Costs

Most of the coordinators state that there are no established practices for assessing costs and allocating budgets for evaluation at their respective departments. For the most part, preliminary budgets are determined once the terms of reference for the assignment have been formulated. Depending on the issues raised here, decisions are made regarding the expertise and time frame required. Sometimes a preliminary time frame is mentioned in the terms of reference, tender documents and the like. The initial assessment to determine whether an evaluation should be done or not – and, if so, to establish its scope – is an informal (and thus for the record invisible) process, distinct from the budgetary process. None of the departments have laid down any formal rules or practices regarding cost ceilings for evaluations. However, an *implicit* assessment of reasonable cost is often made with reference to the size and total cost of the project in question, and, to a certain extent, to the expected value of the evaluation itself.

In cases where an evaluation has been formalized in the agreement with the cooperation partner, financing can be solved in one of several ways. A common practice is to earmark a part of the total project budget for an evaluation or for other forms of follow-up. In such cases, costs are covered from within the project's own budget. More common is that funds are allocated from the IRV (project assessment and evaluation) fund or from other special allowances, that is, from funds that are not project-specific. An exception to this rule is Sida-Öst, where all activities are funded from the same earmarked allowance. Finding funds for evaluations that have been tabled is not seen as a problem.

Dissemination and follow-up

When a first draft of the evaluation report is presented, it is often disseminated among the concerned parties for comment. The comments are incorporated by the consultants in their final draft. This seems to be common practice among all departments. The spread of the report depends to a great extent on its character and scope, but also on the judgement of the responsible programme officer. Project-specific evaluations rarely, if ever, reach beyond a strictly project-oriented readership. Judging by the interviews, no department or unit systematically disseminates evaluations internally for comment or information. Two departments (INEC and NATUR) arrange open, in-house seminars to present and disseminate evaluation results and to provide consultants with feed-back on draft material. Presumably this applies mainly to thematic evaluations of more general scope.

Of great interest is how recommendations and suggestions presented in the evaluations are received and followed up. Recommendations are in all likelihood discussed in the first draft, which, as we have seen, Sida programme officers have the opportunity to comment on. By doing so, however, they are not taking any official position on the recommendations and suggestions presented. Nor are unit chiefs, higher Sida officials or parties concerned in the recipient countries necessarily involved in discussions of such proposals. There are no indications that any department in a systematic manner vets recommendations and suggestions presented in the evaluations. As a consequence, it is difficult to do a systematic *ex post* assessment of the measures taken as a result of the evaluations. In fact, none of the coordinators state that such follow-up is done either at the department or the unit level.

5. The role of the evaluation coordinator

While only a year has passed since the function of the evaluation coordinator was created, it is already relevant to assess its impact. We asked the coordinators what the job has entailed for them so far. Selecting their answers from a multiple-choice format, all coordinators replied that the task has been to disseminate information from UTV within their respective departments, and to collect information within the departments and pass it on to UTV. Four out of nine coordinators answered yes to the question of whether they are kept updated on planned/ongoing evaluations at the department. Most of these coordinators represented small (regional) departments, with very few planned or ongoing evaluations. For coordinators at larger sector departments, it has not been possible to live up to this ambition.

Only one coordinator states that she participates with unit chiefs in the planning and follow-up stages of evaluation efforts. That this is not more common is in all likelihood due to the fact that departmental management, as a rule, does not follow up on evaluation results, either with or without the participation of coordinators (see above). To the extent that evaluations are followed up, it is done at the unit level. All coordinators participated in the half-year and end-year reviews of the 1996 evaluation plan requested by UTV.

Three persons state that they have served as internal consultants/advisors for certain evaluation efforts within their own unit/department. However, they also state that they do not know whether they were approached in their capacities as colleagues, as advisors on other cross-sectoral issue or as evaluation coordinators.

The coordinators believe that the name of the function is known within the departments. However, their actual role has remained unclear to many – to the coordinators as well as to departmental management and other staff. Also, it has been unclear to the coordinators what specific tasks they are expected to perform.

Five out of nine coordinators experience that their respective departments are positive to the new function. The remaining four reply that management is neutral in their attitude towards the function of the coordinators. However, this positive attitude is not reflected in practice. None of the coordinators have been given new work descriptions or more authority as a consequence of their appointment. Four out of nine coordinators participated in all or part of the evaluation training course that was offered in the autumn of 1996.

A problem encountered by all coordinators is the structure and organization of their respective departments. In most of the operational departments, units are completely independent from each other, and there are no natural fora where programme officers from different units can interact. Generally speaking, the role of department management in these units is very limited. So far, department managements have not served as fora for planning and following up on evaluations and their results.

To the question of how they see their own role as coordinators, three out of nine responded that they find the task stimulating. Four reply that the task is burdensome, mainly because the demands have been unclear but also, in some cases, because it has meant an increase in their work load. The latter applies in particular to coordinators for departments numbering several units A majority declares (somewhat surprisingly) that the function is necessary, but most agree that it is unlikely that it would have been created had not UTV requested it.

6. Identified weaknesses in the use of the evaluation tool

In the interviews, the coordinators were requested to identify weaknesses in their respective departments' use of the evaluation tool, and to give suggestions as to how improvements could be made. The respondents found the question interesting and thought-provoking, but were unprepared for it. In the interview form (see Annex 4 to the main report) are listed a number of factors that could be regarded as influencing the use of the evaluation tool. It is to be expected that the responses given previously in the interviews influenced the coordinator's choice of factors to emphasize in response to this question.

The problems raised may be divided in four categories. The first category concerns issues most often raised by the coordinators.[3]

1 Evaluation efforts are not defined on the basis of actual needs. There are no specified goals, nor is there any clear notion of what questions the evaluations are to answer. Also, in many evaluations it is difficult to distinguish the needs of different actors.
2 Evaluation activities are not integrated in a natural and systematic way with the development cooperation efforts. These efforts are fragmented and no one has the broad overview needed for an effective evaluation system.
3 The programme officers' competence in administrating evaluations must be improved. The reports need to be made more accessible, both in terms of form and language and in terms of their availability and distribution.
4 The participation of recipient country actors is weak Sida sets the terms for the evaluations, and cooperation partners are not allowed or encouraged to participate.

It should be noted that it was primarily the coordinators for the regional departments that expressed the need for more systematic coordination of evaluation efforts.

The lack of participation of the cooperation partners was noted by two coordinators with very different experiences in this area from their respective departments.

Many coordinators offered a range of suggestions as to how evaluation efforts could be improved. Some address problem areas mentioned above, others do not. No categories or priorities are made for the listing that follows:

- The purpose of each proposed evaluation must be specified. The issues that are to be addressed should be expressed as precisely as possible.
- Greater effort should be made to involve both Sida and the recipient countries in the evaluations.
- There should be greater participation in the evaluation process.
- Thematic studies should incorporate evaluation results to a greater extent, that is, they should be based on actual experiences and results.

---

[3] Given the multiple-choice format, it was possible for respondents either to indicate a range of problem areas or to limit their replies to one or two specific problem areas This made it difficult to quantify the results further

- Evaluations should be integrated into the country plans.
- All evaluations should be presented with clear and self-explanatory summaries.
- It is important that the evaluations are documented and are readily available, to constitute a link in Sida's institutional memory.
- Programme officers should be offered more advisory assistance in the administration of evaluations.

7. Conclusions

From what has been said above it may be concluded that, as a rule, evaluations at Sida form part of the routine management of individual projects and programmes. There are no formal guidelines as to when or how – or even if at all – an activity should be evaluated. The selection of a project for evaluation is often made at the initiative of the programme officer concerned. There are no formal requirements regarding consultation. In fact, in-house advisers (either within the departments or at POLICY or UTV) are rarely consulted. Sida's own staff are included in evaluation teams only to a very limited degree. Most coordinators expressed the view that persons selected as evaluators should have no ties to the project or programme that he or she is tasked to evaluate.

In a few departments, competitive procurement for evaluation assignments is common Finding suitable evaluation consultants is a common problem. Most – though not all – departments strive to recruit consultants from developing countries.

The evaluators' conclusions and recommendations are disseminated for comment, within Sida mainly on the respective unit level. It is neither the policy nor the practice of any department to take a formal stand on recommendations presented. To date, department managements seem not to have served as fora for planning and following up on evaluations and their results.

The function of departmental evaluation coordinators was created in the beginning of 1996. Half of the first appointees remain, others have been appointed during the course of the year. The main task of the coordinators has been to collect information on behalf of UTV and to disseminate information within their respective departments. In a few cases the coordinator has served as an internal consultant within his/her unit or department. Many have felt their role to be unclear. Also, there seems to be little scope for coordinating this type of activity at the departmental level.

The coordinators identify a range of problem areas in their respective departments' evaluation efforts. Several of these regard limits to requisite competence and capacity within the departments. This needs to be addressed. There is also scope for better identifying the purposes that evaluations may serve with regard to other instruments for following up and analysing Sida's development cooperation.

Appendix: Interview schedule

Department	Evaluation Coordinator	Date of interview
NATUR	Karin Isaksson	12 Nov.
Sida-OST	Maria Lagus	18 Nov.
RESA	Mikael Elofsson	21 Nov.
SAREC	Hellen Olin	27 Nov
POLICY	Berit Rylander	27 Nov.
REWA	Anders Trydell	27 Nov.
INEC	Agneta Danielsson, Leif Holmgren	2 dec
REMA	Samuel Egero	3 Dec.
DESO	Agneta Lindh	6 Dec.
RELA	Elisabeth Hellsten	10 Dec.
SEKA	Lars Bellander	not interviewed
ADM	Lars Boberg	not interviewed
INFO	Lena Tranberg	not interviewed

Annex 3

Summary of interviews with programme officers at Sida

Compiled by Kim Forss and Jerker Carlsson

Table of contents

1. Introduction

The present review presents responses from 30 interview occasions, at which programme officers at Sida were asked about their evaluation experience. The review was initiated at the request of Sida's Department for Evaluation and Internal Audit (UTV), which wished to gain a fuller understanding of how and why evaluations are initiated, who takes part in the evaluation process from the beginning to the end, and how the results are utilized. All evaluations that had been completed in 1996 were selected for the review, and all programme officer responsible for evaluations completed in this period were interviewed.

These 30 interviews fall far short of the 117 evaluations listed in the evaluation plan for 1995/96. However, several of these evaluations were never handed over to the UTV, and it must be concluded that they were never completed. Whether they were postponed, delayed or cancelled is unclear. There is also a small chance that some completed evaluation were never reported to the UTV. The total number of evaluation reports recorded as completed amount to 81 titles for the 1995/96 fiscal year.

The are several reasons why only 30 of these 81 evaluations were included in the study. The interviews were carried out in November and December 1996, and UTV did not receive information from the Sida departments about all evaluations conducted during 1995/96 until January 1997. A few of the 81 evaluations were completed after the interviews were finalised. Furthermore, some programme officers had left the organization prior to the interviews, and it proved impossible to arrange for a suitable time with two of them. Finally, in several instances one person was responsible for more than one report (in which cases he or she was interviewed about one evaluation only). In other instances the evaluations had in fact been completed much earlier, that is, they did not properly belong to the fiscal year 95/96 evaluation plan.

The question is if the fact that the interviewers only reached 37 per cent of the total population has any significance for reliability or validity of the findings. Disregarding the evaluations completed early, could the responses from those programme officers that were not reached be expected to have differed significantly from those that were interviewed? A cautionary reply would be yes, probably in some cases. Without a thorough investigation we cannot ascertain why some evaluations were not reported to the UTV. Nevertheless, we would expect that those in charge of several evaluations would note if anything was highly atypical.

That leaves a shortfall of 6 evaluations that is more troublesome. The fact that some programme officers were unable to fit the interviews into their schedules at any time in almost one month would indicate that they gave the study – and thus possibly the evaluation itself – low priority. Those that left the organization may have done so because they were generally unhappy about working conditions in the wake of Sida's reorganization, or possibly they were pushed out because they did not fit in for some reason or another. There are thus reason to believe that this group would, at least to some extent, have had something different to say about their experiences with the evaluations, and would possibly have given a more negative bias to the responses. Without any systematic investigation we cannot know this, but the results presented below should be reviewed with this in mind.

The interviews followed a structured format (see annex 4). This was established by the Evaluation Secretariat at Sida. Each interview was conducted by one of three researchers (Tove Strömberg, Karin Metell and Lisa Segnestam), who handed over completed printouts of the interviews to the authors of the report (Jerker Carlsson and Kim Forss). There are several weaknesses in such a procedure: the interviews were not conducted in exactly the same way, and the authors missed out on some of the more subtle nuances in the respondents replies that are necessarily lost in a standardized format. We tried to compensate for these inherent weaknesses by arranging two feedback meetings, to share experiences and compare notes during the interview process. Also, one of the authors sat along in one interview with each researcher, to make sure that the process was fairly similar and that the interviews held a high and even standard in terms of reliability. The respondents later had the opportunity to comment on our synthesis of the researchers notes.

This review is meant to be purely descriptive. It summarizes the responses to the interview questions. The analysis, and possibly evaluative comments, are presented in the synthesis report, to which this study is an annex. In theory, the number of responses to any one question should always be 30, but in some instances the respondents could not or would not respond to particular questions. We have chosen to present the remaining responses, well knowing that it can be difficult to correlate these with the responses to the other questions.

2. The respondents

Before turning to the content of the interviews, we need to look at the respondents' background. Out of the 30 evaluations, 7 were produced by INEC, 7 by DESO, 7 by Sida Öst, 5 by SAREC and 4 by NATUR. Table 2.1 presents the background of the respondents. The sums do not add up, as it is typical that people have a diverse working experience. Many have worked in several aid agencies. Those that have worked longest in aid are normally found at SIDA, but others have worked at BITS and the present Sida, or at SAREC, SIDA and again at Sida. It is obvious from the figures that the career patterns are quite disparate in the Swedish aid system. It is notable that none have worked with SwedeCorp.

Table 2.1. Responses to questions regarding which organizations the respondents have a working background in (n=30)

Organisation	Less than 1 year	Between 1 and 2 years	Between 3 and 5 years	Between 6 and 10 years	Between 11 and 20 years	More than 20 years
BITS	1	3	3	-	-	-
SAREC	1	1	1	1	1	-
SIDA	4	1	1	5	2	6
SwedeCorp	-	-	-	-	-	-
Sida	2	28	-	-	-	-
Any other	-	1	6	1	3	-

The last row in table 2.1 indicates whether the respondent has worked for any other organization within the field of development cooperation. Some have for example worked for private consulting firms, others with UN organizations or with NGOs. Some have a background in development research. Somewhat more than half (18) have experience of living in developing countries for more than a year. The remaining 12 have never been in developing countries except on short visits.

The respondents were asked about their experience with evaluations, that is, how frequently they were involved in evaluations tasks. The respondents were given a choice of tasks identified for the purposes of the interviews (see table 2.2,). It is clear that evaluations are a major ingredient in the work of most respondents. All had experience in formulating terms of reference and receiving and following up on evaluations. Only a few had never initiated an evaluation of a project or programme, and only one had never contracted evaluators.

But even though most of the respondents work with evaluations, few have much experience in conducting evaluations themselves. Those who have usually have a background in consulting firms specializing in development and aid, and we assume that these experiences date back to this previous employment.

Given that many have been engaged in evaluation tasks more than four times, it would have been interesting to qualify the question For those who have been working with development cooperation issues for more than 20, four times is not much. But for those who have only been employed at Sida 3–5 years, it is. Although the actual significance of the work load is not clearly established, the indication here is that it is significant.

Table 2 2 *Responses to questions regarding experiences with different aspects of evaluation work, indicating the number of times the respondent has been engaged in each particular task.*

Type of evaluation task	**Never**	**Once**	**Two or three times**	**More than four times**
Initiated an evaluation	4	4	5	17
Written terms of reference for an evaluation	-	6	6	18
Contracted av evaluation team	1	4	4	21
Received an evaluation report	-	2	4	24
Managed the follow-up to an evaluation report	-	4	7	18
Taken part as a member of an evaluation team	21	7	1	1
Taken part as an observer with an evaluation team	21	1	4	3

For easy reference, in the following sections we use subheadings that summarize questions put to the respondents. The purpose is to help the reader locate relevant sections of the text, but also to underscore that the purpose of the text is strictly descriptive, to show what the interviews gave in terms of data.

3. Initiating and planning the evaluation process

Who initiated the evaluation?

The first question in the interview format concerns how and why evaluations are conceived. In 17 cases the respondents were actually the ones who initiated the evaluation. In 3 cases the evaluations were requested by the local embassies. In 4 cases the respondents could not specify how the evaluations were initiated, and in the remaining 6 cases the evaluation was initiated in keeping with requirements set down in the project plan, in other words, the initiative had been taken at the project formulation stage but it could not be said by whom. The main point here is that it is generally the programme officers themselves who initiate the evaluations. It is rare that an external agent, or any other part of the organization, takes the lead.

When was the decision to evaluate taken?

Needless to say, the formal decision to evaluate is always taken by the head of the division, and thus he or she also gets involved at an early stage. The responses to questions regarding how decisions are made generally show that division heads are part of the dialogue at an early stage. The most frequent response to the question of how the evaluation was initiated was: "I came up with the suggestion and then discussed it with the head of the division, who took the formal decision".

It should be noted that the formal decision comes rather late in the process. It is made when money is allocated, and that usually does not happen until the terms of reference are formulated and a budget has been agreed upon with an evaluation team. In other words, the formal decision only confirms a process which by then is already well under way. In one instance the respondent says that the informal decision to evaluate was taken in May 1995, while the formal decision was taken in March 1996.

Why was the evaluation initiated?

There are a number of different explanations as to why the decision to initiate an evaluation of a particular programme or project was taken at all. The most common is that an evaluation has been mentioned earlier in the management of the project. In some cases it is called for in the project document, or in the agreement with other stakeholders. In some divisions it is simply common practice to evaluate a project after a given period of time. In 9 instances this is mentioned as the reason for initiating the evaluation.

Some of the respondents mention that the evaluation object was chosen in order to have something to include in the evaluation plan. The activity selected either seemed "suitable" or the programme officer thought it would be "interesting" to have this particular activity included in the evaluation plan. In this connection, it was often mentioned that the selected activity had never been evaluated before, and that it thus made good sense to include it in the evaluation plan. These slightly overlapping reasons were also mentioned in 9 instances, where the need for something to include in the evaluation plan was cited by 4 respondents and the wish "to know about results" was cited by 5 respondents.

Another reason given was that the project had come to an end, or was approaching a mid-term review. In both cases, Sida would soon be entering negotiations on whether and how the programme should continue. Here an evaluation would contribute to the decision-making process. This is mentioned as the reasons to evaluate by 4 respondents.

3 evaluations were initiated because the division in question was expected to make a contribution to the development of country strategies. The recent reorganization of the aid agencies has also influenced the choice of evaluation objects. In two cases the evaluations were initiated to document former BITS projects and that organization's methods of working with aid.

However, several interviews do not give any clear response to the question, and often the responses seem to overlap. An evaluation was perhaps initiated to generate information needed for the work on a country strategy *and* because it was required by agreement with other stakeholders. There could of course be other combinations of motives.

Who were consulted when the evaluation was initiated?

The initiative to evaluate was usually not taken in isolation. To what extent were other stakeholders consulted when the decision to evaluate was taken? Responses are presented in table 3.1. Note that in theory each category of stakeholder could be represented 30 times, but in practice the range of contacts is far lower. The most frequent contacts are with the programme officers' immediate superiors: heads of the divisions or subdivisions. A total of 82 stakeholder consultations are mentioned in the interviews, which means on average roughly three per evaluation.

Table 3 1 Consultations with other stakeholders regarding the decision to evaluate

Stakeholder category	Number of respondents indicating they had a contact with someone in this stakeholder category:
Evaluation coordinator	1
Head of division/department	24
Colleagues in the division	9
Evaluation Secretariat	5
Other units at Sida	4
Embassies	9
Recipient country organisations	12
Project consultants	9
Other consultants	6
Other donor agencies	3

What was the purpose of the evaluation?

The interview format returns to the issue two questions later, when the respondent is asked about the purpose of the evaluation. But here the question is formulated more directly: Did you have any notion of the purpose of the evaluation, and of how its results were to be used, at the time when the decision to evaluate was taken? The responses to this question are far more comprehensive and explicit. In fact, many interviews quote the objectives that are stated in the terms of reference for the evaluation. Bellow follow some examples of what the programme officers said in respect to different objectives.

First and most important is immediate managerial use. This is mentioned as the primary purpose of the evaluations of, among others, the following projects: research cooperation on tropical rain forest management systems; cooperation between Swedish county administration boards and the Baltic countries; production in north-western Russia; support to the road sector in Estonia/Latvia/Lithuania; a biotechnology project; and the opening of two road corridors in Angola. All of 17 respondents mentioned managerial use as the most important reasons for the evaluation, as the following responses serve to illustrate:

> We intended to use the evaluation results in future decisions on whether (this consulting firm) could be used in the future.

> The purpose was to assess the programme and give recommendations for future support. This could be used in planning, rectify problems, determine how much money would be needed, and improve the planning.

> We were uncertain about the effects of the programme, and what results that had been achieved. The project had lasted for several years, and there was now a proposal to continue, and thus we needed an evaluation

Several of the evaluations were expected to yield criteria for success that could be used in future decisions to approve – or reject – projects, as for example the evaluation of concessionary credit schemes in Zimbabwe and the evaluation of Swedish support to the forestry sector in Latvia. (The evaluation of BITS support to telecommunications is a similar example. In this case the respondent states that the purpose was to learn about development effects of support in this sector, while at the same time developing methods to assess project proposals in the sector). Nine of the respondents mention such consideration as the most important reason for the evaluation. Others mention – often in very vague terms – such transfer of experience as being of only secondary or tertiary importance.

Some programmes had a pilot character, and it was therefore deemed necessary to analyse results and study implications for other programmes, as well as for the future of the particular project. One such example is the evaluation of SIDA/SAREC's marine science programme. The evaluation of the strategic business alliances in Costa Rica is another. Here three programmes were evaluated in parallel, and the purpose was to compare approaches and distil recommendations for future programme support.

Finally, a few respondents identified their evaluations in terms of more long-range learning objectives. One such example is the above-mentioned evaluation of concessionary credits schemes in Zimbabwe. The evaluation covered 10 concessionary credits and was expected to show a pattern of successful – or failed – investments. It was expected that the lessons of this programme would be of use when taking decisions on future credits to Zimbabwe. The evaluation of the Sri Lankan–Swedish research cooperation programme also addressed both broader issues regarding the effectiveness of SAREC support and the need for background information for upcoming bilateral negotiations regarding the future of the programme.

A small minority of respondents – 4 in all – cite a need to know project/programme results as the first and foremost reason for the evaluation.

To summarize, of three broad categories of evaluation needs – audit, management and learning – an overwhelming majority of the respondents identify management needs as the prime motor for the evaluations, followed by learning needs. Only very few evaluations were initiated to satisfy audit needs.

How were the terms of reference established?

We should remember that the terms of reference of the evaluations are usually formulated after consultations with others at Sida. The respondents were asked how widespread these contacts were. The answers were recorded in the same way as in table 3.1 above. These results are presented in Table 3.2.

It is clear that such contacts are still few and far between. While unit or division heads are the most frequent contacts, it is perhaps surprising that not all respondents had contact with these immediate superiors. Generally speaking, respondents had more contacts within Sida in preparing the terms of reference than in initiating the evaluation.

Only 12 of 30 had any contacts with recipient government organizations. There were also few that sought contact with other donors or organizations. None sought contact with their respective evaluation coordinator.

Table 3.2. Contacts during the preparations of terms of reference

Stakeholder category	Number of respondents indicating they had a contact with someone in this stakeholder category:
Evaluation coordinator	0
Head of division/department	23
Colleagues in the division	15
Evaluation Secretariat	12
Other units at Sida	9
Recipient country organisations	12
Project consultants	0
The evaluation team	6
Other donor agencies	7
Other Swedish agencies	2

This raises the question of whether consultations with other stakeholders influenced the terms of reference or in any other way resulted in changes in the direction of the evaluation. One of the evaluations is reported to have been changed considerably in the light of consultations, 18 were not changed, and for the remaining 11 there were minor changes. These were of two kinds it was suggested either that some specific item should be added – notably a discussion of project/programme cost-efficiency – or that the evaluation should be better focused or better structured. The former comments mainly came from UTV, the latter often arose out of contacts with division heads or with other units at Sida.

Were reference or steering groups used?

The next question concerns the organization of a reference or steering group to monitor the evaluation process. In only 2 of the 30 interviews the respondents confirm that reference or a steering groups had been set up. 2 others state that such groups were planned but were for different reasons never set up. 5 respondents mention that they worked closely with colleagues or with their unit heads, and so in effect formed informal reference groups.

Which background documents were used to formulate the terms?

All respondents use common project documents, such as quarterly reports, annual reports, other progress reports or special reports from consultants, previous evaluations (if any) and the project or programme documents. All respondents consulted these sources.

The majority had read, and referred to, Sida's evaluation manual. They are asked whether they knew about Sida's evaluation policy, and most did, but many were not aware of it at the time of writing the terms of reference. A majority are satisfied that they had found the documents needed, most of which were available at their respective divisions. Four respondents mention problems in finding background documents, and one could not answer whether the necessary documents were found. In one case the respondent explains that as the evaluation consultants wrote the terms of reference, the respondent did not know whether it was difficult to find the documents.

How and when were the costs of the evaluation calculated?

The respondents were asked about the cost estimation of the evaluation. 24 of the respondents said that they made a rough estimate based on the approximate number of working weeks that would be needed, the extent of travel and a ball-park figure for the fee usually charged by an outside consultant. Two respondents state that there was a ceiling for evaluation costs, but they do not know how that ceiling was calculated. Two respondents state that the evaluation was put on tender and that the consultants had to suggest a budget. Price then became one of the selection criteria. In both cases no budget estimates had been made prior to the tendering process. Finally, two respondents did not know how the costs were calculated.

The evaluations were financed from the IRV budget (10 evaluations), the country frames (7), the budget heading for contract-based technical cooperation (7), a separate budget line for environmental studies (2) and in one case from a regional budget line. None of the respondents indicated that the evaluations were insufficiently financed. One even felt that the evaluation was over financed. (In 3 cases we could not determine how the evaluation had been financed.)

Were the assignments put on competitive bidding?

In 25 cases, the evaluator or evaluation team was contacted directly. Only 5 evaluations were put on tender. The most common reasons for why not more were put on tender were that the budgets were too small, there were so few persons that qualified, or a tender procedure was administratively cumbersome. Several respondents mention that the cost of the evaluation was below 200,000 – alternatively 300, 000 – so there was no need to do a tender. One respondent reported that the tendering procedure was evaded by splitting up the total costs of the evaluation. Tendering is required by Swedish law and Sida also has a policy for procurement of goods and services.

How were the evaluators identified?

The evaluation teams, or the evaluators, were mainly identified through informal personal contacts. The programme officers asked colleagues in the division, people at the embassy or even outside organizations. Many of the selected evaluators were already well-known to the respondents (15 cases). Twice as many respondents claim that it is easy to locate consultants for evaluations than claim that it is difficult. Only one respondent had turned to UTV's consultancy register to find names.

What qualifications were needed for the evaluation assignment?

The type of competence requested by the respondents varies, but most mention sector competence and knowledge of specific subjects as prime considerations. Many also looked for someone with a knowledge of the country, the region and the local language. Other valued qualifications are evaluation experience, knowledge of development issues and knowledge of cross-cutting issues (gender, the environment, democracy and human rights, free market reforms, etc). Skills in financial analysis are called for the least, and are mentioned by only two respondents.

Of the 30 evaluation teams, 6 include local consultants. A few respondents indicate that they had wished to include local evaluators but that for various reasons this was not possible. Of the remaining 20 respondents, many never considered the idea and most were sceptical, arguing that it would have been impossible to find someone with the necessary skills.

How much time was spent on the evaluation?

It is during the evaluation of the tenders that the programme officers discuss issues pertaining to the design of the evaluation. Once this has been agreed on, questions regarding methods and theory do not form part of the programme officers involvement in the planning of the evaluation. This does not mean that they are inactive or leave the evaluators on their own. On the contrary, they are very active. The most common participation pattern can be described as "paving the way". They assist evaluators in tracking down background information from agency archives and elsewhere.

They also assist in contacting other concerned parties, in the recipient country and in Sweden, to inform them of the coming evaluation and to ask them to cooperate with the evaluator. This involvement means that an evaluation can often be quite time-consuming for the programme officer. Table 3.3 gives an idea of the magnitude of work involved. The distribution is fairly even across the range of alternatives, and no particular trend can be discerned. It is noteworthy that as many as 8 respondents spent more than 15 days working on an evaluation. A word of caution is called for here since time reports are not used at Sida, the respondents cannot always be precise in their estimates of time spent working on the evaluations.

Table 3 3 Working days spent on an evaluation by the programme officer

Number of working days	1-5	6-10	11-15	16-20	More than 20	Not applicable no answer
Number of respondents	6	4	5	4	4	7

4. Implementing the evaluation

To what extent was the evaluation plan changed during implementation?
During the practical planning phase, the evaluation continues to be a collegial affair. The programme officers often consult with colleagues, particularly within his or her department. But it was rare for any of them to suggest a different approach to the evaluation. Such comments, when made, seldom concern issues of method or other technical aspects. In one case a department head wanted to ensure that the evaluation did not concern itself too much with the role of programme officer responsible for the project. In another instance one of the regional departments wanted an earlier completion date, in order to use the evaluation in the reporting of project results.

What is the role of the programme officer during the conduct of the evaluation?
It is less common for the programme officer to take an active part once the evaluation has started. But in a few cases the programme officer continuously discussed the progress of the evaluation with the evaluator. The type of contacts vary from regular telephone briefings to informal meetings. None of the respondents report that they had used a more organized procedure to channel feedback during the implementation stage, for instance through reference groups where the evaluator presents his results for comments during the course of the work.

How were draft reports circulated?
In our survey we gave much attention to the report itself. The respondents were asked several related questions about their assessment of the evaluation. However, the programme officers' assessment of the quality of the report is not something that can be analysed at only one point in time. Once data collection is completed and the first analyses are presented, there are a series of stages involving comments and revisions of the draft report before a final text is ready. The programme officers' reactions/assessment of the report can therefore be expected to change over time. In order to capture this process we tried to focus on two points in time – when the first draft arrives and when the final report is presented.

When the first draft arrives, it is circulated among individuals and organizations concerned. For the purpose of the interviews, we identified 11 different categories of persons and institutions. The programme officers were then asked to choose from this list the categories of persons to whom they sent the report for comment, and to indicate which of these actually submitted any comments. Table 4.1 reveals an interesting pattern.

Table 4.1. Who were asked to comment on the first draft of the evaluation report?

	Was asked	Responded
Evaluation coordinator	0	0
Division head	22	18
Colleagues in the department	13	12
Evaluation Secretariat	10	3
Other Sida departments	13	7
Swedish embassy/DCO	11	9
Recipient country authorities	14	10
Local project staff	6	5
Target groups	0	0
Implementation consultants	18	14
Any other	2	2

The programme officers solicit the opinions of colleagues and other stakeholders on the first draft. Some reply. Some are asked more frequently than others about their views. The most important reference is the head of the department. 73 per cent of the respondents sent the draft to the departmental head. The consultants responsible for the implementation of the project was the second most frequent reference (60 per cent). In roughly half of the cases, the evaluation was sent to the institution in the recipient country responsible for the project. The programme officers' colleagues in the department are sometimes (in 43 per cent of all cases) asked to give their views. Colleagues in other departments at Sida were consulted with equal frequency.

The Swedish embassy in the receiving country, and the evaluation unit in Sida, are asked to comment in one third of all cases. Judging by the responses, it is uncommon for the programme officers to solicit comments from the local project staff, but it is possible that these are covered by the category "implementation consultants". Not unexpectedly, the ultimate beneficiaries of the support are never asked to comment on the evaluation. Finally, the evaluation coordinator at the department level is not consulted in any of the cases, presumably because the function was not yet established at the time when most of the evaluations were being conducted.

Discussions of the draft report

Do the stakeholders respond with comments? The response ratio forms a simple measure of how actively they are involved in the evaluation. Colleagues in the department almost always give comments . So do the Swedish embassies in the collaborating country, department heads and local project staff. At the other end of the spectrum we find UTV, which rarely comments on draft evaluation reports.

The way comments are submitted varies considerably. A distinct pattern could not be found in the responses. Comments are either mailed or communicated verbally to evaluators and/or programme officers. A common procedure is to arrange meetings between evaluator and the programme officers, where the latter has the opportunity to convey views of other stakeholders. Sometimes programme officers arrange internal meetings with colleagues in his/her department. It is less common to arrange open seminars where reports can be discussed among the concerned stakeholders.

What do the programme officers think of the first draft?

How do the programme officers react to the first draft? Almost all of them are critical (83 per cent). Generally, the critique focuses on formal aspects such as poor language, unclear outline and vaguely formulated conclusions and recommendations. While several respondents criticise the analysis for lack of depth, they are not prepared to challenge the type of analysis used or the facts and figures presented by the evaluator. The criticism more often reflects dissatisfaction with the results.

How were comments received?

The evaluators were receptive to the programme officers' comments on the drafts. The respondents felt that their comments and suggestions were normally taken into account. Most respondents felt that the final report was, by and large, true to the intentions of the terms of reference. However, some expressed disappointment regarding some aspects, usually the analysis of impact and cost-effectiveness. In instances where the terms of reference specifically called for an analysis of impact and cost-effectiveness, the evaluators had encountered problems that had resulted in a weak analysis. A broad picture of the programme officers' assessment of the final report is given in table 4.2. below.

Table 4 2 Satisfaction/dissatisfaction with the final evaluation report

Is the final report satisfactory in respect of:	Yes	No	Cannot tell
Language, clarity and organisation	27	3	0
Methods and data sources	26	4	0
Reliability	27	2	1
Objectivity	26	2	2
Practical usefulness	20	8	2
Wealth of information	17	13	0
New ideas	20	8	2

How is the final report assessed?

The picture of the evaluation process that emerges here is quite clear. The programme officers are generally satisfied with the final reports, which are deemed to be well-written and well-structured. The information is easily accessible. The evaluations almost always present the methods applied and data sources used in such a way that it is clear to the reader how the results have been obtained. The results are also felt to be very reliable The evaluations also give a balanced and objective view of what has happened in the evaluated projects. The interests and views of the various stakeholders are represented in a fair way

However, the programme officers are less enthusiastic with the practical usefulness of the recommendations of the evaluation, although a majority feels that they are clear and implementable. They also feel that there is a logical link between the analysis of project/programme results and the recommendations made. However, when asked if the evaluation provided new knowledge about the evaluated activity, some 40 per cent responded negatively. The programme officers seemed to possess a fairly deep knowledge of the projects, to which the evaluations often had little to add. On the up side, when asked if the evaluations presented new ideas and a new perspective on the evaluated activity, about two thirds of the respondents acknowledged that this was the case. The evaluation also stimulated them to reflect on other issues in the area of development cooperation.

In order to shed more light on their assessment of the evaluation, the respondents were asked to elaborate on what they saw as the strengths and weaknesses of the evaluation. The replies vary considerably, and it is difficult to aggregate the answers into distinct categories. There is no common pattern. What is regarded as a strength in one evaluation may be seen as a weakness in another. However, by way of crude generalization, most respondents identify the following characteristics as strengths: a project-specific orientation, a concentration of analytical efforts on project achievements, and a perspective that places the project in a country context. A weakness identified by several respondents is a failure to assess the impact of the project on the surrounding society.

5. Impact and use of the evaluation

One of the major purposes of an evaluation is to serve as an input to the decision-making process. This leads to questions about how the findings of the evaluation are consumed by the various stakeholders and other interested parties. The pattern of use found in the survey is summarized in table 5.1.

Is the quality of the report uncontested, or did it cause controversy and conflict among the stakeholders? It seems that in about half the cases the reports were accepted by all parties. It is interesting to note that in 40 per cent of the cases the respondents did not know how other stakeholders had reacted to the final report.

The final report is read by those directly concerned by the project. To what extent is it distributed to persons outside this group? In about half of the cases the evaluation was given wider circulation. Most respondents felt that the reports had been widely read and discussed. To what extent other parties shared the programme officers assessment of the quality of the evaluation is less clear. In about half of the cases they were not aware of any differing opinions.

Table 5 1 Dissemination and discussion of the evaluation

	Yes	No	Do not know
Report quality undisputed?	14	4	12
Report distributed to external parties?	14	13	3
Widely read and discussed?	18	8	4

Are the evaluations practical and useful?

First and foremost, the evaluations provide an analysis of the performance of the project. This was in itself regarded as important and useful by the programme officers. But equally interesting is of course whether the evaluation had any direct practical implications. Did it result in any immediate and concrete actions? Table 5.2 summarizes responses to questions in this regard.

Table 5 2 Were the recommendations accepted and applied?

	Yes	No	Do not know
Were the recommendations accepted?	17	6	7
Did they result in concrete action?	22	5	3
Is there a likelihood of any future actions?	12	11	7
Actions other than those recommended?	7	16	7

Notwithstanding some dissatisfaction, the evaluators recommendations were clearly accepted in about half of the cases. It is noteworthy that in 6 cases the client did not approve of the recommendations. In as many as 7 cases the respondent did not know. In 73 per cent of the cases it was possible to translate the recommendations into concrete action. The following excerpts serve to illustrate the range of practical measures that were taken in response to the recommendations.

> The evaluation concluded that the collaboration between the Indian researchers did not work In the future, SAREC will therefore give support to only one institution The evaluation also recommended continued support along the same line

The reference group has been given a more active role. The programme officer is now trying to integrate the marine projects with the rest of Sida (the coastal initiative). Phase II will be implemented.

One project will receive support for another three years. The other project will be terminated within two years. The evaluation stresses the importance of accounting for results and recipient ownership, which are two recommendations taken on board by the programme officer.

Sometimes it is not the recommendations that lead to action, but the report as a whole:

One sub-project was given a much smaller budget allocation due to the severe criticism in the report (although not a direct consequence of the recommendations) The evaluation is used to a large extent by the ministry when formulating the new agreement.

It is difficult to be specific about the likelihood of any future action. Suffice to say that one third of the respondents feel that the evaluation will continue to have a future impact. Perhaps more interesting is that just as many are certain that the evaluation will not result in any future actions. Finally, on a few occasions actions were taken that were not recommended in the evaluations.

Lessons learned in the evaluation process

The respondents were asked to reflect on lessons learned in the evaluation exercise. One of the most frequently cited lessons was that evaluations take time. Not only in terms of the time a programme officer needs to spend on the evaluation, but also in terms of the duration of the whole evaluation process, from start to finish. None took less time than 3 months, 11 took 3–6 months, another 11 took 7–12 months, 3 took 13–18 months, and 5 took longer than 18 months.

It seems it is difficult to conduct an evaluation in less than three months. In 73 per cent of the cases the whole process took up to a year. In five extraordinary cases it took up to two years. Obviously such lengthy procedures were not intended. Despite careful planning, the evaluations were for various reasons fraught with delays. Some had to do with finalising of the report – language problems and other editorial matters. In some cases it took an unduly long time to have questionnaires filled in and returned. In other instances it took a long time to find all the relevant background material. Delays that occurred in the course of information gathering sometimes upset the evaluators time schedules, which may have provoked further delays later on.

In some cases the programme officer experienced problems with the evaluator. At issue was not necessarily the technical qualifications of the evaluator, but just as often the working relationship between the programme officer and the evaluator. Irrespective of the causes to the problem, it increased the respondents understanding of the need for careful procedures in selecting evaluators.

Tendering procedures – assessment of tenders, and dialogue with the different bidders as part of the assessment – was an issue that the respondents felt they needed to pay greater attention to. At the root of many of the problems encountered with evaluators, and of the degree of satisfaction with the final report, was the terms of reference for the evaluation. An important experience for many respondents was the recognition of the need to write terms of reference and to formulate questions that are realistic and researchable, given the constraints of the evaluation. Many respondents noted that next time they will ask the UTV for guidance.

Appendix: Interviewed programme officers with respective evaluation

Andersson, Ingvar
Valdelin, J. *et al.*, Rural *Village Water Supply Programme – Botswana*, 96/10.

Bergstrom, Jonas
Markstedt, A., *Support to the Road Sector 1993–1996 in Estonia, Latvia and Lithuania*, 96/31–33.

Bhagavan, Malur
Schügerl, K., *Biotechnology Project: Applied Biocatalysis*, 96/4.

Broman, Angelica
Kenny, L. and Petren, A., *Sida Support to Environmental Awareness and Traning Projects through the Panos Institute, Gemini News Service and Television Trust for the Environment*, 96/41.

Burman, Fredrik
Goppers, K., *Concessionary Credits in Support of Economic Development in Zimbabwe*, 69/42.

Bålöw, Ros-Mari
Neito, A., *Sida/SAREC Support to Collaborative Programme for Biomedical Research Training in Central America*, 96/19

Gerhart, Karin
Whitmore, T. C., *Swedish–Malaysian Research Cooperation on Tropical Rain Forest Management*,96/18.

Granlund, Anders
Rudegren, J. *et al.*, *Sida/SAREC's Marine Science Cooperation Programs*, 96/35.

Gyllenhammar, Kerstin
Boström, K., *Swedish Support to the Foresty Sector in Latvia*, 96/15.

Hedlund, Anders
Svenningsson, P. L., *The Baltic Sea* (not published).

Holmgren, Leif
Eriksson, C. H. and Moller, G., *The Opening of Two Road Corridors in Angola*, 96/37.

Hultberg, Marianne
Grenstedt, L., *Statistikproduktion i Nordvastra Ryssland*, 96/38.

Irani, Farhad
Price Waterhouse KB and Galaxicon AB, *Evalution of Consulting Services within the Planta Nicaragua Rehabilitation Programme* (not published).

Johansson, Bengt
Valdelin, J. *et al.*, *UNICEF's programme for water and sanitation in Central America: Facing new challenges and opportunities*, 96/11.

Johansson, Jan
Lindqvist, M. *et al.*, *Konvertering av rysk militärindustri*, 96/8

Liljesson, Lars
Rylander, L. *et al.*, *Telecommunications: A Swedish Contribution to Development*, 96/3.

Löfgren, Åke
Heldahl, I. A. and Hoffman, J., *Evalution of Sweden's Support to Mayebuye Centre, University of Western Cape and Appraisal of Swedish Support in the Future* (not published).

Nunes-Sörensson, Eva
Ballara, M. *et al.*, *Avaliacão do Apoio Sueco ao Sector da Educação na Guiné Bissau 1992–1996*, 96/7.

Persson, Hans
Pehrsson, K., *Educação Ambiental em Moçambique*, 95/1.

Rehlen, Christina
Almqvist, L. C. G., *Cooperation between the Swedish County Administration Boards and the Baltic Countries*, 96/17.

Rosencrantz, Kerstin
Bäckman, O. *et al.*, *Evalution Study of the Special Education Programme in Zimbabwe* (not published).

Sher, Afzal
Sande ratne, N. and Nilsson, J. S., *Sri Lankan–Swedish Research Cooperation*, 96/39.

Sundelin, Dag
Andreasson, B. *et al.*, *The Electricity Sector in Mozambique, Support to the Sector by Norway and Sweden*, 96/21.

Sundgren, Margareta
Olsson, B., *Water Supply in Dodota – Ethiopia*, 96/23.

Tanttari, Liisa
Falk, H. and Wallberg, B., *Svenskt stod till vänortssamarbete med Polen, Estland, Lettland och Litauen*, 96/22.

Tullberg, Margareta
Oppelstrup, H., *National CBR in Zimbabwe* (not published).

Wedekull, Christina
Helander, M., *Strategic Business Alliances in Costa Rica*, 96/28.

Werner, Gösta
Green, B. and Law, P., *Swedish Support to Botswana Railways*, 96/16.

Wickman, Kenth and Sundberg, Emma
Vadelin, J. *et al.*, *Sida Support to the Education Sector in Ethiopia*, 1992–1995, 96/27.

Arnfast, Inger
Alberts, T. and Alexandersson, C., *The Swedish Fisheries Programme in Guinea Bissau*, 1977–1995, 96/20.

Annex 4

Guideline forms for interviews with evaluation coordinators and programme officers

Interview with evaluation coordinators

No.

Background and purpose (see separate document)

Name of person interviewed:
Unit/Department:
Interviewer:
Date of interview:

1. Background information on the evaluation coordinator

1.1 How many years have you worked in the development cooperation field?

At BITS years
At SAREC years
At SIDA years
At Swedecorp years
At Sida
Other aid organization

1.2 How many years/months have you worked in the area that you are currently working in?

1.3 Have you ever lived and worked in the field? Yes No

What were your tasks?
For what organization?
For how long? years

1.4 How many months have you been evaluation coordinator? months

Approx. 9 months since the function was created

1.5 What is your previous experience with evaluation work?
Have you:

	Never	1 time	2–3 times	4–
Initiated an evaluation?				
Been responsible for the formulation of the terms of reference?				
Tendered bids from consultants and researchers for evaluation assignments?				
Received and commented an evaluation report?				
Followed up on an evaluation's recommendations and suggestions?				
Participated in the evaluation field work?				
- as member of an evaluation team?				
- as an observer?				

2. Departmental structures and policies for evaluation The purpose of this section is to profile those formal departmental policies and/or guidelines that include guidelines for evaluations conducted within the department's scope of activities. It is also intended to highlight how evaluation work is conducted in practise, that is, the normal procedures followed in the planning, implementation and follow-up of the department's evaluations.		
2.1 Are there any written guidelines on evaluation work specific to the department?	Yes	No
2.2 Are there any clear directives on evaluation work in other sector-specific policy or strategy documents?	Yes	No
Specify:		
2.3 Are there any formal rules or standard procedures to determine what is/is not to be evaluated? This could involve criteria such as size of support, duration and date, the evaluation of programmes as opposed to projects (or vice versa), trouble-shooting, preparations for renewal of agreement, compliance with agreement, etc.	Yes	No
Specify:		

2.4 a) Who in your unit decides whether an activity should be evaluated? Chief of Unit
Head of Department

2.4 b) Who usually initiates evaluations? Thirteen (13) potential actors are listed below. Discuss their respective roles in the initiation and authorization of evaluations.

Comments:

1. Programme officer
2. Chief of Unit
3. Head of Department
4. Evaluation coordinator
5. Adviser
6. Regional department
7. Sector department
8. Embassy programme officer
9. Embassy counsellor
10. Cooperation partner in recipient country
11. Government authority in recipient country
13. Other

2.5 Describe the standard operating procedure for administrating evaluations. What form(s) of cooperation take place within and between the department? (Describe (a) the role of different actors, (b) how common the cooperation is, and (c) what form(s)it takes. Use the above list of potential actors.)

2.6 Are there any norms or rules regarding who should/should not be entrusted with evaluation tasks (for instance, consultants in recipient countries, Swedish/foreign consultants, one-man/multiple staff teams, Sida's own staff, project staff, persons with special expertise, etc.)?

2.7 (Estimate) To what extent is competitive bidding used in contracting evaluation consultants?

Comments:

2.8 Is an internal register of consultants available for reference?	Yes	No
2.9 a) Are there any rules or established practises regarding evaluation cost limits?	Yes	No
How is it determined what is a reasonable cost for an evaluation?		

2.9 b) How, as a rule, are the department's evaluations financed? (funding source)

2.10 a) Are there any rules or established practices regarding cooperation with other donors? In what form(s) do(es) such cooperation take place at (a) planning, (b) implementation and (c) dissemination of evaluation results?

2.11 a) Are there any rules or established practices regarding the dissemination and follow-up of evaluation results?

Ex. Seminars are arranged to discuss the report	
Drafts of the report are distributed for comments and views	
At the follow-up of recommendations	
Other	

4. What does the role of evaluation coordinator (EC) imply?			
4.1 What has the job of evaluation coordinator entailed so far?			
a) The EC collects and/or disseminates information at the request of UTV.			
b) The EC is kept updated about planned/ongoing evaluations at the department.			
c) The EC participates in department management's planning of coming evaluation efforts and in the follow-up of completed evaluations.			
d) The EC serves as internal consultant/adviser for certain evaluation efforts within his/her own unit/department.			
e) Other			
4.2 How would you describe the attitude of the management of your department towards the role of the evaluation coordinator?	Positive	Negative	Neutral
How does this manifest itself in practise? (time, resources, influence, training, responsibility, etc.)			
4.3 How do you yourself regard your role as evaluation coordinator. Do you feel the job is . . . onerous, stimulating, ungratifying, appreciated, essential, marginalized, etc. Feel free to use your own words.			

5. In what way(s) could the department's evaluation work be improved?
What weaknesses can the evaluation coordinator identify in the department's evaluation work? How could the evaluation programme be improved? In particular, how could the evaluations better be put to use?
(Examples of factors that *may* influence the way evaluations are used are:
availability, quality, timing, participation of cooperation partners, greater independence, diffusion, improved competence of responsible programme officers, etc.)
Ask question first, await reply, then list suggested areas

6. In the view of the evaluation coordinator, what should a central administration for evaluations at Sida (UTV) focus more attention on?

	Priority
Advisory support (i. e., assistance in formulating terms of reference, assess bids, provide consultant registers, assess incoming reports)	
Education/training	
Assess and follow up departments' evaluations (with regard to quality and application)	
General methodological issues (providing indicators, manuals, etc.)	
Generating its own (independent) evaluations If so, what kind of evaluations?	
Other	

Comments:

Interviewer's notes
Time: minutes
Respondent's involvement in the interview:
Reliability of interview:

Interview with programme officers

No.

Title (and no.) of evaluation report
Name of person interviewed:
Unit/Department:
Interviewer:
Date of interview:

1. Background information		
1.1 How many years have you worked in the development cooperation field?		
	At BITS	years
	At SAREC	years
	At SIDA	years
	At Swedecorp	years
	At Sida	years
	Other aid organization	years

1.2 Have you ever lived and worked in the field?	Yes	No

1.3 What is your previous experience with evaluation work?
Have you:

	Never	1 time	2–3 times	4–
Initiated an evaluation?				
Been responsible for the formulation of the terms of reference?				
Tendered bids from consultants and researchers for evaluation assignments?				
Received and commented an evaluation report?				
Followed up on an evaluation's recommendations and suggestions?				
Participated in the evaluation field work?				
- as member of an evaluation team?				
- as an observer?				

2. Planning and preparations

You've answered a few general questions about your background. We now turn to the evaluation that you administrated as responsible case officer.

2.1 When, by whom and why was the initiative taken to conduct the evaluation?

2.2 Who was contacted in connection with this initiative? *(Allow first the respondent to give a spontaneous reply, then follow up)*	**Contact:**
a) The department's evaluation coordinator b) The head of the department/unit c) Other colleagues at the department/unit d) UTV e) Other Sida departments f) Embassy in recipient country g) Recipient country government authorities local project staff target groups h) Implementing consultants i) Other consultants j) Other donor organizations	
2.3 When was evaluation authorized and by whom?	

2.4 When authorization was given, was there a clear notion as to the purpose of the evaluation and to how the evaluation would be used?

2.5 Was a steering or reference group set up for the evaluation? Why? How was it created? Who did it include? How did it work?	Yes	No

2.6 Who did you contact when you wrote the terms of reference? *(Allow first the respondent to give a spontaneous reply, then follow up)*	**Contact:**
a) The department's evaluation coordinator b) The head of the department/unit c) Other colleagues at the department/unit d) UTV e) Other Sida departments f) Embassy in recipient country g) Recipient country government authorities local project organization target groups h) Other donor organizations i) Other Swedish authorities j) Implementing consultants k) The consultants contracted to do the evaluation	

2.7 How where the terms of reference changed as a result of these consultations?

2.8 On what material did you base your draft of the terms of reference?
a) Background documents for the project

b) Earlier evaluations of the same project/programme
c) Evaluations of other projects/programmes
d) Methodological manuals
e) Sida's evaluation policy (are you familiar with it?)
f) Other documents

2.9 Did you find the documents you needed? Account for any difficulties you encountered	Yes	No

2.10 How were the expected costs for the evaluation calculated?		
2.11 From what budget post was the evaluation funded *(IRV, FOM, country frame/project funds)*		
2.12 In your view, were sufficient funds allocated for the evaluation. If not, explain why.	Yes	No

2.13 What type of expertise was of special value to the task?
a) Regional expertise, b) sector-specific expertise, c) expertise in the subject matter, d) evaluation competence, e) special competence in analysing gender, environment or other cross-cutting issues.
(Refer first to ToR, then follow up if necessary.)

2.14 How were the evaluation consultants identified? (Consultant register, authors of previous evaluations suggestions from colleagues, etc.) Was it difficult to find suitable consultants?

2.15 Did you require or express a wish that the evaluation team included a member from the recipient country? Why/why not? If so, where did this wish originate?

2.16 Was competitive bidding used to select consultants? Why/why not? How many bids were made?

2.17 Who participated in the selection of bids (recipient country, embassy, UTV, evaluation coordinator, etc.)? What were the primary selection criteria?

3. Implementation
Now we move from the evaluation planning and preparation stage to the implementation stage.

3.1 Did anyone have a differing view regarding how the evaluation should be carried out (choice of consultants, timing, purpose, cost, formulation of ToR)? Account briefly for differing views.	Yes	No

3.2 How did you follow the evaluation work from the signing of the contract to the presentation of the first draft of the report?

3.3 How long did it take from the fist initiative to the final report?

3.4 Did it take longer than expected? If so, why? Did the delay affect the use of the report?	Yes	No

3.5 Who was asked to comment on the draft? Who commented, and how (in writing/in conversation)?	Asked:	Commented:
a) The department's evaluation coordinator b) The head of the department/unit c) Other colleagues at the department/unit d) UTV e) Other Sida departments f) Embassy in recipient country g) Recipient country government authorities local project staff target groups h) Implementing consultants i) Others		

3.6 How was the first draft disseminated? Were seminars and/or other oral presentations arranged? Other channels?

3.7 Did you have critical comments to the draft? If so, what were they and how did you express them (in writing/in conversation)? If not, why?	Yes	No

3.8 Was the text of the final report changed as a result of comments received?	Yes	No

3.9 Does the content of the final report correspond to the terms of reference? If not, what are your main objections in this regard?	Yes	No

3.10 In your view, what are the strongest points of the evaluation report?

3.11 In your view, what are the weakest points of the evaluation report?

3.12 Does the finished report satisfy your expectations as regard to:		
a) Language, style and disposition. Is the presentation clear and comprehensible?	Yes	No
b) Discussion of method. Are sources and methods for data collection and analysis presented in a satisfactory fashion?	Yes	No
c) Reliability. Do you consider the report's findings reliable?	Yes	No
d) Comprehensive and impartial with regard to the interested parties. Does the report present a fair and balanced view of the evaluated activity? Is due consideration given to the views and interests of all parties concerned?	Yes	No
e) Practical guidance. Does the evaluation present the parties concerned with clear and practical recommendations regarding feasible measures? Are the recommendations clearly motivated by the preceding analysis?	Yes	No
f) Wealth of information. Has the evaluation report provided you with new, essential information regarding the evaluated activity? Have you learned something unexpected about the evaluated activity?	Yes	No
g) Wealth of ideas. Does the report open new perspectives on the evaluated activity? Does it stimulate ideas of relevance to other cooperation support activities?	Yes	No

Comments:

3.13 Do other interested parties share your views on the quality of the report? Give examples of any differences in view.	Yes	No	Don't know

3.14 Has the report been spread outside the circle of parties directly interested in the project? Have you received comments from these persons? If so, what are they?	Yes	No	Don't know
3.15 Has the report been read and discussed to the extent that you wished? If not, why not?	Yes	No	

3.16 Have the recommendations in the evaluation report been accepted by the parties concerned?	Yes	No
3.17 Has the evaluation resulted in any concrete measures in keeping with the recommendations? Which?	Yes	No
3.18 Is there any reason to believe that the evaluation will result in further measures being taken in the future? Which?	Yes	No

3.19 Has the evaluation resulted in any measures other than those recommended in the report?	Yes	No

3.20 How much time do you estimate you've spent on the evaluation?
3.21 Have you learned anything about the evaluation process that you feel should be shared with others?

Background and purpose of interviews with Sida's evaluation coordinators

UTV is conducting a study of Sida evaluations commissioned and completed in the course of 1996. The study, entitled "Using the evaluation tool: A survey of conventional wisdom and practice at Sida", focuses mainly on how and why evaluations are initiated; how planning, authorization and evaluation consultant procurement is managed; and how the results are used by Sida. The study is to make recommendations on possible improvements regarding evaluation planning and implementation, and the feed-back of evaluation results. The basis for this follow-up is some forty-odd interviews with programme officers managing evaluations commissioned by Sida's various departments in 1996.

This interview is one of some ten-odd interviews with Sida's evaluation coordinators. Our main interest is to gain an overall perspective on how evaluation work is organized within the departments, and on the rules or norms that guide these activities. We raise the issue of regulations and practises with the aim of gaining a descriptive – if not complete or comprehensive – picture of the respective departments' evaluation efforts. The function of the evaluation coordinator is recent, and some of the respondents have only limited experience of evaluation work at other units within his/her own department.

The information gleaned from both sets of interviews will be collated to provide a basis for comparisons of how the different departments make use of the evaluation tool. The aim is to highlight both weaknesses and positive experiences. Our hope is that people from all parts of Sida will learn from each others' experiences in working with the evaluation tool.

Annex 5

Quality Assessment of 48 Sida Evaluations
Compiled by Tove Strömberg, SIPU International

Table of contents

1. Introduction

This annex is part of a report that deals with the evaluation process within the Sida organization. The project team has used three different methods for gathering information:

1 Quality assessment of 48 evaluation reports registered in 1995–96;
2 Interviews with 30 Sida programme officers responsible for at least one evaluation report in 1995–96;
3 Interviews with 10 evaluation co-ordinators within the Sida organization.

This annex describes the result of the quality assessment of the 48 evaluation reports. The standardized worksheet and the methods used are described in the main report. The interviews with the Sida programme officers concerned 30 of the 48 evaluation reports mentioned in this appendix.

Table 1 1 Number of evaluations assessed, by originating department

Department	Number of evaluations
Department for Central and Eastern Europe	13
Department for Democracy and Social Development	10
Department for East and West Africa	2
Department for Infrastructure and Economic Co-operation	9
Department for Natural Resources and the Environment	6
Department for Research Co-operation/SAREC	6
Evaluations produced in co-operation between departments	2
Total	**48**

As table 1.1 shows, the Department for Central and Eastern Europe is the most common producer of evaluation reports in our sample. The departments for Democracy and Social Development and for Infrastructure and Economic Co-operation are also frequent producers. The departments for Administration including Sida-Sando, Information, Policy and Legal Services, Evaluation and Internal Audit, Asia, Southern Africa and Latin America are not represented in this sample of evaluations. Two departments have produced evaluations in co-operation with other departments. The Department for Democracy and Social Development has produced an evaluation in co-operation with the Embassy of Sweden in Ethiopia. The Department for Co-operation with Non Governmental Organizations and Humanitarian Assistance has produced an evaluation in co-operation with the Embassy of Sweden in Angola and with the Department for Infrastructure and Economic Co-operation.

2. Characteristics of the evaluation objects

This section contains a description of what kind of support the 48 reports evaluate. The following items are described:

- Type of support
- Region supported
- Money used in the project during the evaluated period
- Time period
- Type of sector
- Aid channel
- Recipient organization

According to table 2.1, over 80 per cent of the evaluations in the sample describe support to individual projects or project-implementing organizations. 15 per cent evaluates support to several linked projects within the same sector and with one common set of objectives. 4 per cent evaluates general financial support given under special conditions, i.e., balance of payment support. The same pattern is repeated, if not so pronounced, in the sample of 30 evaluations where interviews followed.[1]

Table 2 1 Characteristics of the evaluation object

Type of support	Project	Sector	Programme aid		
Total 48	81%	15%	4%		
Total 30	73%	20%	7%		

Region	Africa	Asia	Latin America	Europe	Global
Total 47	47%	8%	13%	30%	2%
Total 29	45%	10%	14%	28%	3%

Evaluated earlier	>1	1	0	Unknown	
Total 47	17%	21%	38%	23%	
Total 30	20%	23%	33%	23%	

Aid channel	Multilateral	Bilateral	NGO		
Total 48	6%	92%	2%		
Total 30	7%	90%	3%		

[1] In all tables, all percentages are rounded to the nearest whole integer, which means that totals may not add up to equal 100 per cent In some cases it was not possible to extract the relevant information from either the evaluations or the Sida Evaluations Data Worksheets (see annex 6). As a consequence, the number of evaluations cited in the tables is not always 48 or 30.

Recipient	Public	Mixed	Private	NGO
Total 48	87%	6%	4%	2%
Total 30	90%	0%	7%	3%

Almost half of the reports evaluate support given to countries in Africa. 30 per cent evaluate support to Eastern and Central Europe, and 13 per cent and 8 per cent evaluate Latin America and Asia, respectively. Two evaluations could not be assigned to any given region. One evaluates the total support given to the telecommunication sector and one evaluates the support given to the Panos Institute Gemini News Service and Television Trust.

Almost 40 per cent of the evaluated projects have never been evaluated before. This information would have been hard to gather without the Sida Evaluations Data Worksheet (see annex 6[2]) as it is not provided in most evaluations. For this reason, this variable is unknown in all of 23 per cent of the evaluations.

94 per cent of both the larger sample of 48 evaluations and the smaller sample of 30 evaluate support provided through bilateral aid channels. In most cases, the recipient is a public authority.

Table 2 2 Size of support (MSEK) compared to duration of support (years)

Years	MSEK < 1	1–10	11–50	> 50	Total
< 2	1	2	1	0	4
2–5	1	5	11	4	21
6–10	0	0	6	2	8
> 10	0	0	1	5	6
Total	**2**	**7**	**19**	**11**	**39/39**

Table 2.2 compares information from the evaluation reports regarding the size and duration of the support. In some cases information for one or both variables was missing. The table provides data for the 39 evaluations for which the relevant information is available. Almost one third of these reports evaluate support of more than MSEK 50 distributed in the course of the evaluated time period. Half the reports evaluate support to a value of MSEK 11–50. Of the 48 evaluations, only 40 reported project costs. The evaluators have stated the amount of support Sida has provided in the Sida Evaluations Data Worksheet. If this information had not been given through the worksheet it would have been difficult to extract it from the evaluations. None of the evaluations tried to assess what the actual total cost for the project was. The total cost includes not only Sida's support but also the support given by the recipient country or by other organizations. The reported support varied substantially. One evaluation involved a project receiving Sida support to a value of MSEK 0.155, another involved a project receiving support totalling MSEK 5000

More than 60 per cent of the support was evaluated after less than five years. The reason for this high number is the high percentage of reports evaluating Eastern and Central Europe. None of the reports produced from the Department

2 The Sida Evaluations Data Worksheets were completed by the authors of the evaluations conducted for Sida. The Worksheets detail information of relevance to the evaluations, such as originating department, evaluation cost, project cost, project type, etc

for Central and Eastern Europe evaluated a longer time period than five years. Many of these are old BITS projects, which usually were short-term and did not involve a substantial amount of money.

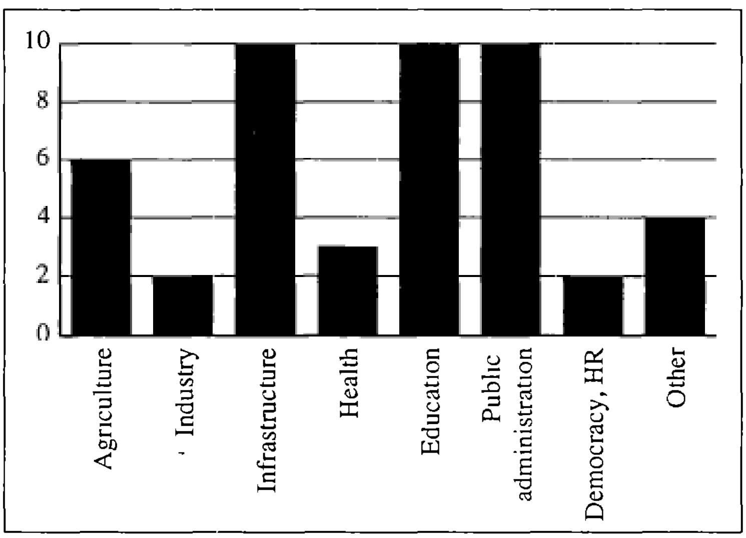
For these variables, there is no significant difference between the total sample of 48 evaluations and the smaller sample of 30 evaluations. If the samples are broken down by region, however, some differences appear. In both samples, most evaluations of support over MSEK 50 regard projects in Africa. In the larger sample, Africa also ranks highest for evaluations of support in the range 11–50 MSEK. But in the smaller sample, Eastern and Central Europe ranks highest for evaluations of support in this range.

As shown in figure 2.1, three development sectors rank equally high in the larger sample, each being the target for more than 20 per cent of the evaluated support.

1 Support to *infrastructure*, including all support to the development of transportation (railways, roads, etc.), communications (telecommunications, postal services, radio, etc.) and energy production.
2 Support to *education*, including primary- and secondary education and support to universities and research institutions.
3 Support to *public administration*, including public financial administration and banking, central and local administration and support to privatization and the development of labour markets, trade unions and legal systems.

In the smaller sample, this sectoral break-down looks slightly different. Infrastructure still accounts for slightly more than 20 per cent, but education increases its share to 24 per cent and public administration shrinks to 14 per cent. However, all three remain the largest sectors.

Figure 2 1 Frequency distribution at sector



To summarize, most of the studied reports evaluate bilateral development support to Africa in excess of MSEK 10. The support is concentrated to the sectors infrastructure, education and public administration and has been ongoing for less than five years.

3. The timing and cost of the evaluation process, and the composition of the evaluation team

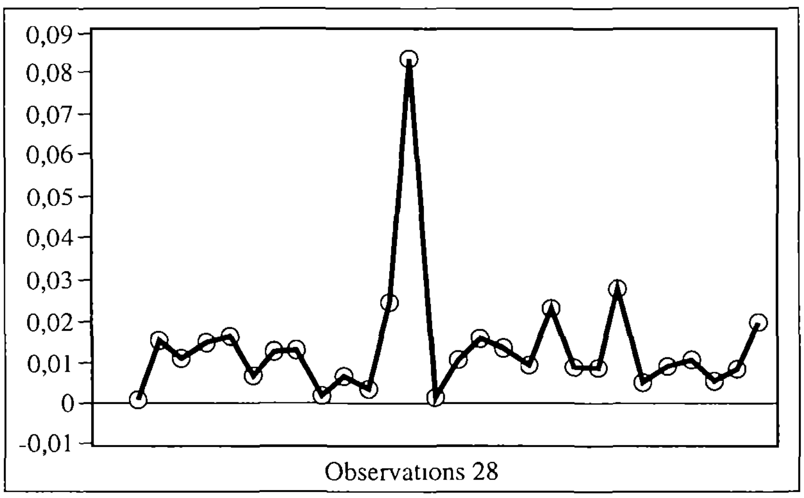
This section describes the timing of the evaluation, the costs to Sida of the evaluations in the samples, and the composition of the evaluation teams.

In both samples, more than half of the evaluations are made mid-term. Approximately 30 per cent are made at the time of completion of the project. In the total sample, 15 per cent are made ex-post and in the smaller sample 21 per cent are made ex-post. In all cases, ex-post evaluations are made two years or more after completion of the project.

All but five evaluations are financed by Sida only. Four evaluations are co-financed by Sida and either Orgut, the World Bank, Norad or the Swiss Development Cooperation Agency. One evaluation was financed by the former Swedish Board for Investment and Technical Support (BITS).

It has been possible to identify Sida's cost for the evaluation in 37 cases. These figures show that on average the evaluations cost around 370 000 SEK. Among the 30 evaluations interviewed on the cost can only be identified in 25 cases, but the average is roughly the same as in the total sample.

Figure 3.1 Evaluation cost / Sida support



A comparison of the costs to Sida of the evaluations and the value of Sida's project support during the periods under evaluation shows that, in the total sample, the evaluation cost corresponds, on average, to 1.3 per cent of Sida's total investment in the project during the period in question. This figure decreases to 1 per cent in the smaller sample. These figures vary considerably depending on the nature of the support (see figure 3.1[3]) The minimum amount spent on any one evaluation was 0.03 per cent of the total investment, and the maximum amount spent was 2.3 per cent. No correlation can be shown between the costs of the

[3] It some cases, there was insufficient data to extract information regarding either the cost of the evaluation or the value of Sida project support during the period in question The figure is based on the 39 evaluations for which such data was available.

evaluation and the value of Sida's support to the projects during the periods in question. In the samples, costs for evaluations of projects receiving less than MSEK 50 in Sida support range between MSEK 0.1 and MSEK 1.

The number of persons in the evaluation teams vary from one to seven. In approximately 40 per cent of all evaluations, the teams consist of just one person. Women participate in approximately 30 per cent of the teams. Two of the evaluations are written only by women.

Table 3 1 Composition of evaluation teams

	Yes	No	Unknown
Agency/donor present			
Total 48	0%	94%	6%
Total 30	0%	93%	7%
Recipient country/regions present			
Total 48	27%	67%	6%
Total 30	20%	73%	7%
Sector expert competence present			
Total 48	83%	4%	12%
Total 30	80%	3%	17%

The information in table 3.1 about the evaluation teams' composition was rarely available in the evaluations, and was therefore in most cases extracted from the Sida Evaluations Data Worksheets. In some cases the information was extrapolated by the persons who studied the evaluations.[4] Neither Sida nor other donors are represented in the evaluation teams. A representative from the recipient country is part of the evaluation team and co-authors of the report in 27 per cent of all cases. The most common evaluator is a person with a sector competence

4 Karin Metell Lisa Segnestam, Tove Stromberg

4. Characteristics of the evaluation report

This section describes the structure of the reports, with an eye on:

- type of evaluation[5]
- reporting language (English, Swedish or language of recipient country)
- appendices included

These variables affect both the utility and the propriety of the report. As table 4.1 shows, most evaluations are project evaluations, reflecting the fact that most evaluation objects are support to projects. Around 20 per cent of the evaluations study series of projects linked through a common set of objectives to form a programme. None of the evaluations address all support given to a country or an organization. The samples include few sector, programme aid or thematic evaluations.

Table 4 1. Type of evaluation and reporting language

Type of evaluation	Project	Pro-gramme	Sector	Program-me aid	Them atic	Coun-try	Organi-zational	Other
Total 48	71%	19%	4%	2%	2%	0%	0%	2%
Total 30	63%	20%	7%	3%	3%	0%	0%	3%

Reporting language	Swedish	English	Beneficiary country language
Total 48	4%	88%	8%
Total 29	7%	86%	7%

English is the most common report language. Two evaluations are written in Swedish and four in the recipient country's language, which in the samples are Portuguese and Spanish.

Most evaluations include both a summary and the terms of reference for the evaluation (see table 4.2). Only half of the evaluations include recommendations to the recipient country, while almost all evaluations include recommendations to

[5] Sida evaluations can be classified under eight different type headings·
1 Project evaluations address a limited number of activities with a common set of objectives, or a set of projects of similar kind not forming a programme from the donor's point of view
2. Programme evaluations address a series of linked projects with a common set of objectives
3 Sector evaluations address all, or the majority of, Sida-financed projects for the sector.
4 Country evaluations address all, or the majority of, Sida-financed projects and programmes in a country where development sectoral objectives and/or Sida's country strategy or the eqvivalent forms the basis
5 Programme aid evaluations focus on financial programme support, i e , balance of payment support (import support and debt relief) and general or sectoral untied budget support given under specific policy conditions
6 Thematic evaluations focus on thematic issues such as gender support to Vietnam or telecommunications support in Africa.
7. Organizational evaluations address an organization financed fully or partly by Sida Focus is on organisational capacity, not primarily on the effects of the project
8 Other other type of evaluation study or review.

Sida or to other donors. In some cases, the information was extracted from the Sida Evaluations Data Worksheets or extrapolated by the persons who studied the evaluations. The data imply that most evaluators regard Sida or other donors, and not the recipient country, as their target group.

The questionnaires or interview formats have been annexed in 12 per cent of the evaluations. Around 40 per cent of the evaluations present the raw data.

Table 4.2. Summary, terms of reference, recommendations, questionnaire and raw data

	Yes	No
Summary included		
Total 48	90%	10%
Total 30	90%	10%
Terms of Reference included		
Total 48	89%	11%
Total 29	93%	7%
Recommendations to be pursued by donor included		
Total 48	90%	10%
Total 30	90%	10%
Recommendations to be pursued by recipient included		
Total 48	52%	48%
Total 30	43%	57%
Questionnaire or interview formats annexed		
Total 48	12%	88%
Total 30	17%	83%
Presentation of raw data		
Total 48	38%	62%
Total 30	43%	57%

To summarize, most evaluations in the samples are project evaluations, are written in English and include a summary and the terms of reference for the evaluation.

5. Methodology and presentation of methodology

This section describes the methods used and how the evaluators present them in the reports.

Table 5.1 demonstrates to what extent the various methods for data collection have been used in relation to the evaluation tasks. Only a few of the evaluations discuss the sources used or the functions of the persons interviewed. Typically, the discussion of methodology is limited to a brief statement along the following lines:

> The evaluation is based on documentation of the programmes, scientific reports produced by the programmes, and extensive discussions with persons concerned.

The evaluations assessed in this survey use and describe methods for data collection minimally and or not at all. There is no difference between the total sample and the sample of 30 evaluations that were accompanied with interviews. This makes it difficult to assess the use of various methods for data collection. Where lists of the persons interviewed are provided, they usually only specify names and organization affiliation. It is difficult to assess the value of a piece of information gleaned in an interview if the status of the interviewed person is not known. The most common category of people identified in the evaluations as sources of information is project staff.

In the smaller sample of 30 evaluations, 27 per cent of the evaluations were assessed to have used and described in an exemplary fashion interviews with persons from the recipient country. The corresponding figure for the larger sample of 48 evaluations is only 16 per cent.

Most evaluators visited the recipient country to observe the project work. As a rule they visited the authorities and organizations involved; only rarely did they visit the final beneficiaries.

Table 5 1 Data collection

	Exemplary treatment	Addressed adequately	Addressed minimally	Not addressed
Agency document used				
Total 48	10%	12%	52%	25%
Total 30	13%	10%	53%	23%
Use of other documents				
Total 48	14%	19%	38%	29%
Total 30	17%	23%	33%	27%
Agency personnel interviewed				
Total 48	14%	17%	25%	44%
Total 30	20%	17%	23%	40%
Project staff interviewed				
Total 48	17%	47%	30%	6%
Total 30	20%	57%	20%	3%
Recipient interviewed				
Total 48	17%	14%	42%	27%
Total 30	27%	7%	37%	30%

Other donor interviewed				
Total 48	2%	14%	14%	69%
Total 30	3%	10%	13%	73%
Direct observation of project work				
Total 48	2%	8%	69%	21%
Total 30	3%	3%	77%	17%

The use of protocols or guides is scarce. When used they are usually not annexed to the report. Only 13 per cent of the evaluations used and annexed a questionnaire.

In 50 per cent of all evaluations the question of "participatory evaluation methods" is not addressed; only 11 per cent make use of them. ("Participatory evaluation methods" means that the evaluator has visited and talked to the recipients in the country). An example is when an evaluation team has visited several villages in a water project. None of the evaluations have used such formal techniques as gauging cost–benefits or economic rate of return.

This lack of accounts with regard to documentation, interviews, protocols, guides or questionnaires makes it difficult to judge the accuracy of the evaluations. This is exacerbated by the fact that the evaluators' hypotheses and assumptions are adequately stated in only 6 per cent of the evaluations (see table 5.2). Typically, the presentation of hypothesis and assumptions is limited to brief statements along the following lines:

> Our hypothesis is that all evaluations include adequate presentation of their methodology.

or

> We assume that an adequate presentation of the methodology will lead to evaluation reports that have a high accuracy.

The assessment of the variable "hypothesis and assumptions of the..." includes reviewing definitions of words and expressions. The assessment shows that the use of definitions such as the following are scarce

> the main aim of the support is to transfer administration, professional knowledge and experience to strengthen the administrative capability and democratic process In other words, the predominate part of the approach is institution building

Table 5.2 Methodology

	Exemplary treatment	Addressed adequately	Addressed minimally	Not addressed
Methodology issues are addressed				
Total 48	2%	21%	64%	12%
Total 30	0%	23%	63%	13%
Hypothesis and assumptions of the evaluation team are made explicit				
Total 48	0%	6%	17%	77%
Total 30	0%	3%	27%	70%
Discussions of validity of results				
Total 48	0%	4%	12%	83%
Total 30	0%	3%	13%	83%
Discussions of reliability of results				
Total 48	2%	6%	19%	73%
Total 30	3%	7%	23%	67%

Few evaluations discuss issues of validity or reliability, that is, they do not discuss if the evaluations answer the questions stated in the terms of reference and/or if they have used the right methodology to answer these questions. Overall, the evaluations assessed in this survey address methodological issues only to a minimal extent. The only difference between the larger and the smaller sample is that the larger sample includes one evaluation that gives exemplary treatment to methodology.

6. Report analyses

This section assesses the analyses made in the reports with regard to support objectives, effectiveness and efficiency.

In assessing to what extent support objectives are discussed in the evaluations, the following questions are relevant:

- Does the evaluation describe the objectives of the support?
- Do the evaluations state whether the objectives are specific, measurable, realistic and have a time limit?
- Do the evaluations discuss whether the support objectives are relevant to the project?
- Do the evaluations discuss whether the objectives are met?

Table 6 1. Analyses of the main objectives of the support

	Exemplary treatment	Addressed adequately	Addressed minimally	Not addressed
Development objectives				
Total 48	0%	19%	36%	45%
Total 30	0%	23%	43%	33%
Immediate objective achieved				
Total 48	17%	40%	29%	14%
Total 30	17%	30%	40%	13%
Key project assumptions				
Total 48	0%	10%	40%	50%
Total 30	0%	10%	50%	40%

As table 6.1 shows, many of the evaluations address the development objective minimally or not at all. However, the two samples differ somewhat: In the smaller sample, more evaluations address the development objective minimally than not at all. In the larger sample the opposite is true. Among the evaluations that were made ex-post more than 40 per cent address the development objectives adequately, while among the evaluations made mid-term the figure is only 12 per cent.

Several of the evaluators identify the lack of relevant project documentation and monitoring reports as a problem This may be one reason why as many as a third of the evaluations treat immediate objectives only minimally. More evaluations give adequate attention to immediate objectives in the larger sample than in the smaller sample.

Effectiveness is a measurement of the extent to which a project or programme is successful in achieving its objectives. The lack of discussion of objectives makes it hard to discuss effectiveness. Only half of the studied evaluations discussed effectiveness adequately and none did so in an exemplary fashion. Nearly 20 per cent of the evaluations did not address effectiveness at all.

Table 6.2. Other analysis made in the reports

	Exemplary treatment	Addressed adequately	Addressed minimally	Not addressed
Country's need for the project				
Total 47	0%	28%	45%	28%
Total 30	0%	30%	43%	27%
Comparison with/without the project				
Total 48	2%	4%	19%	75%
Total 30	0%	3%	20%	77%
Project expenditure analysis				
Total 48	0%	27%	54%	19%
Total 30	0%	20%	63%	17%
Project benefit analysis				
Total 48	6%	42%	38%	14%
Total 30	3%	43%	43%	10%
Assessment efficiency				
Total 48	2%	17%	38%	44%
Total 30	0%	13%	43%	43%
Discussion of aid agency's performance				
Total 48	2%	14%	48%	35%
Total 30	3%	17%	50%	30%
Discussion of general lessons				
Total 48	2%	17%	38%	44%
Total 30	3%	17%	43%	37%

The variable "key project assumptions" (see table 6.1) pertains to the question of whether the evaluation identifies any external factors beyond the control of project management, affecting the objectives, output, input as well as impact of the support. Some evaluations include a description of external factors, but none discuss their impact on the support. Half of the evaluations in the larger sample do not addressed this issue, while half of the evaluations in the smaller addresses the issue minimally. Only 10 per cent in both samples treat the question adequately.

The variable "country's need for the project" is given similar treatment in the reports. Many evaluations describe the recipient country's economy, political situation, etc. The implication is that the status of the country makes the support relevant, but in none of the evaluations is the linkage made explicit. Another reason why none of the evaluations are assessed to have made exemplary analyses of the recipient countries' needs is that none discuss why Sweden is relevant as a donor country. There is also a lack of analysis of opportunity costs, that is, the costs if no support had been given.

Analyses of expenditure and efficiency (the relationship between costs and benefits) show the same deficiencies as the analyses of objectives and opportunity costs. According to the evaluators, also here there is a lack of documentation and monitoring. Not one of the evaluations was able to identify the expenditure of the recipient country. The evaluations also lack a discussion of indirect costs, such as environmental costs accrued in the building of roads. As table 6.2 shows, less than 20 per cent of the evaluations do not discuss project expenditure at all. For obvious reasons, no assessment of cost-effectiveness is possible in these instances.

More than 80 per cent of the evaluations discuss the benefits of the support. Evaluations made at the end of the support or ex-post more often address this issue adequately than evaluations made mid-term. The reason is probably that it is only possible to estimate the impact of a support in a long-term perspective.

Most evaluations are project-oriented and do not monitor how the support from the donor organization has been administrated. More than one third of the evaluations does not address the question at all, and almost half address the performance of the aid agency only minimally. Again, there are no differences between the two samples.

It is more common for the evaluations to give recommendations than to discuss what general lessons can be drawn from the project. In the larger sample, 44 per cent do not address the question at all. This is more than in the smaller sample, where 37 per cent do not address the question and 43 per cent address the question minimally. All of the evaluations made ex-post discuss general lessons in some way. None of the evaluations made by completion of the support or mid-term address the issue in an exemplary way.

7. Project sustainability

This section discusses how the evaluations treat the question of sustainability, that is, the factors influencing sustainable impact of the support. For example, evaluating support to personnel training schemes includes looking at questions like:

- How many of the persons that have attended training remain within the organization?
- Does the organization have the means to continue training personnel after support is terminated?

Some 70 per cent of the evaluations in both samples address sustainability adequately or minimally (see table 7.1). The lack of discussion of expenditure has an impact on the discussion of sustainability. More than half of the evaluations do not discuss alternative financing, and 60 per cent do not discuss financing from earned revenues. The most common sustainability discussion is the commitment of the recipient country to sustain the gains.

Table 7 1 Analyses of sustainability

	Exemplary treatment	Addressed adequately	Addressed minimally	Not addressed
Sustainability of project after agency withdrawal				
Total 48	2%	31%	40%	27%
Total 30	3%	23%	47%	27%
Alternative financing of project/programme after agency withdrawal				
Total 48	2%	10%	33%	54%
Total 30	3%	3%	33%	60%
Degree of financing of project/programme from earned revenues				
Total 48	4%	17%	19%	60%
Total 30	3%	17%	20%	60%
Commitment of recipient country/local management for sustainability				
Total 48	2%	23%	48%	27%
Total 30	3%	20%	40%	37%

Often the reason behind a mid-term evaluation is that Sida wants a basis for taking a decision about further support. Since more than half of the evaluations are made mid-term, most evaluations discuss if it is motivated to continue the support and not if the support is sustainable. Among the evaluations made ex-post, more than half have an adequate or exemplary treatment of sustainability. Among evaluations made mid-term, 75 per cent address sustainability minimally or not at all. The trend is the same for both samples.

8. Cross-cutting issues

This section discusses the treatment of issues of general character, such as gender, the environment, human rights, democracy and market economy.

As seen in table 8.1, the most frequently discussed cross-cutting issues are gender and the environment. Both gender and environment are mostly minimally addressed.

An example of such a discussion is the following:

> We are told that these trucks have a proven record of being more environment friendly than many of its competitors in that it has lower levels of emission. The contrast to the fleet trucks operating before the purchase of the new trucks is said to be dramatic. Whereas the older trucks were constantly embedded in clouds of foul-smelling exhaust fumes, the new trucks are said to have attracted attention in the street because no one could see the exhaust fumes So, we venture to conclude that the purchase of the trucks did in fact contribute to the air becoming somewhat cleaner

The smaller sample address gender adequately or minimally more often than the larger sample. Only two evaluations treated gender issues in an exemplary fashion. Issues of human rights, democracy and market economy are almost never addressed.

Table 8 1. Analyses of cross-cutting issues

	Exemplary treatment	Addressed adequately	Addressed minimally	Not addressed
Gender				
Total 48	4%	8%	42%	46%
Total 30	3%	10%	50%	37%
The environment				
Total 48	0%	10%	31%	58%
Total 30	0%	13%	37%	50%
Human rights				
Total 47	0%	2%	13%	85%
Total 30	0%	3%	13%	83%
Democracy				
Total 48	0%	4%	19%	77%
Total 30	0%	7%	20%	73%
Market economy				
Total 48	0%	0%	10%	90%
Total 30	0%	0%	10%	90%

9. General assessment of the sampled evaluations

Every evaluation has been rated according to an overall assessment of all performance variables. This rating provides a general assessment of each evaluation. The relative importance of each variable tends to vary depending on the evaluation. In other words, "exemplary" performance is not required for all variables discussed above for an evaluation to be rated "excellent".

Table 9 1. General assessment of sampled evaluations

Assessment	Excellent	Good	Adequate	Inadequate
Total 48	6%	27%	48%	19%
Total 30	3%	27%	53%	17%

Almost half of the evaluations in both samples are assessed as "adequate". Even if the evaluations lack some analysis and a description of methods, they still reach reasonable conclusions. Most evaluations have a good structure and include terms of reference, summary and recommendations. Most are written in English. Evaluations of Sida support seem to be conducted in a regular fashion, usually within two to five years after the project start.

One reason that so few evaluations are regarded as "excellent" or "good" is that they often lack a discussion of methodology, including data collection. Another reason is that they often lack thorough analysis of objective fulfilment, impact, sustainability, cross-cutting issues and project expenditure.

The samples are too small to draw any general conclusions regarding differences between ex-post, mid-term and completion evaluations, or regarding evaluations made by evaluation teams including or excluding women. However, both these areas show interesting patterns.

In our small samples, ex-post evaluations tend to have a more thorough discussion of objective fulfilment, impact, sustainability and general lessons.

Evaluations made by evaluation teams including women tend to rate higher overall than evaluations without women. Two of the three evaluations that were regarded as "excellent" were conducted by teams that included women. Of the 13 evaluations that were regarded as "good", more than half included a woman in the evaluation team. 9 evaluations were regarded as inadequate; none of these had a woman on the evaluation team.

Both these variables should be further analysed using the whole data base, since the sample it too small to draw any general conclusions.

Appendix: Evaluations assessed

(* = selected for inteview, see annex 3)

Almqvist, L. C. G., *Cooperation between the Swedish County Administration Boards and the Baltic Countries*, 96/17.*

Alberts, T. and Alexandersson, C., *The Swedish Fisheries Programme in Guinea Bissau*, 1977–1995, 96/20.*

Andersson Brolin, L. and Silvia, M., *Programa de Vivienda Social de FUPROVI – Costa Rica* (not published).

Ballara, M. *et al.*, *Avaliacão do Apoio Sueco ao Sector da Educação na Guiné Bissau 1992–1996*, 96/7.*

Bendz, M., *Forest Sector Development Programme, Lithuania–Sweden*, 96/13.

Boström, K., *Swedish Support to the Foresty Sector in Latvia*, 96/15.*

Brook, I., *Cadastral and Mapping Support to the Land Reform Programme in Estonia, Lithuania and Latvia*, vol. 1, 96/24.

Brook, I. and Ragnar, C., *Cadastral and Mapping Support to the Land Reform Programme in Estonia, Lithuania and Latvia*, vol. 2, 96/29–30.

Bäckman, Ö. *et al* , *Evalution Study of the Special Education Programme in Zimbabwe* (not published).*

Sande ratne, N. and Nilsson, J. S., *Sri Lankan–Swedish Research Cooperation*, 96/39.*

Andreasson, B. *et al* , *The Electricity Sector in Mozambique, Support to the Sector by Norway and Sweden*, 96/21.*

Davey, R., *Botswana Road Safty Improvment Projects*, 96/43.

Eriksson, C. H. and Moller, G., *The Opening of Two Road Corridors in Angola*, 96/37.*

Falk, H. and Wallberg, B., *Svenskt stod till vänortssamarbete med Polen, Estland, Lettland och Litauen*, 96/22.*

Goppers, K , *Concessionary Credits in Support of Economic Development in Zimbabwe*, 69/42.*

Green, B. and Law, P., *Swedish Support to Botswana Railways*, 96/16.*

Grenstedt, L., *Statistikproduktion i Nordvästra Ryssland*, 96/38.*

Heileman, A. and Peck, L., *The Beira-Gothenburg Twinning Programme*, 96/1.

Heldahl, I. A. and Hoffman, J., *Evalution of Sweden's Support to Mayebuye Centre, University of Western Cape and Appraisal of Swedish Support in the Future* (not published).*

Helander, M., *Strategic Business Alliances in Costa Rica*, 96/28.*

Karlén, I., *Social Information System, Estonia* (not published).

Kamugisha, J. R. and Semu, E., *Soil and Water Conservation Research Project at Kari, Muguga – Kenya*, 96/26.

Kenny, L. and Petren, A., *Sida Support to Environmental Awareness and Traning Projects through the Panos Institute, Gemini News Service and Television Trust for the Environment*, 96/41.*

Lindqvist, M. *et al.*, *Konvertering av rysk militarindustri*, 96/8.*

Malmqvist, T. and Wallenberg, B., *Cooperative Environment Programme: Asian Institute of Technology/Sida, 1993–1996*, 96/12.

Markstedt, A., *Support to the Road Sector 1993–1996 in Estonia, Latvia and Lithuania*, 96/31–33.*

Nabais, J. *et al.*, *Estructuração do Sistema Nacional de Gestão de Recursos Humanos*, 96/6.

Nars, K., *Debt Management (Kenya)*, 96/2.

Nieto, A., *Sida/SAREC Support to Collaborative Programme for Biomedical Research Training in Central America*, 96/19.*

Palme, M*et al.*, *Curriculum Development in Ethiopia, A Consultancy Study for the Ministry of Education In Ethiopia and for Sida*, 96/40.

Pehrsson, K., *Educação Ambiental em Moçambique*, 95/1.*

Price Waterhouse KB and Galaxicon AB, *Evalution of Consulting Services within the Planta Nicaragua Rehabilitation Programme* (not published).*

Olsson, B., *Water Supply in Dodota – Ethiopia*, 96/23.*

Oppelstrup, H., *National CBR in Zimbabwe* (not published).*

Rudegren, J. *et al.*, *Sida/SAREC's Marine Science Cooperation Programs*, 96/35.*

Rylander, L. *et al*, *Telecommunications: A Swedish Contribution to Development*, 96/3.*

Schugerl, K., *Biotechnology Project: Applied Biocatalysis*, 96/4.*

Skåring, M., *Evalution of the Ongoing Project: Construction and Implementation of an Accounting and Financial System for the Estonian Public Administration* (not published).

Svenningson, P. J., *Support to the Development of Civil Aviation Administration in the Baltic States*, 96/36.

Svenningsson, P. J., *The Baltic Sea* (not published).*

Thulstrup, E. W. *et al.*, *Building Research Capacity in Ethiopia*, 96/9.

Tiffen, M. *et al.*, *National Soil and Water Conservation Programme: Kenya*, 96/25.

Valdelin, J. *et al.*, *UNICEF's programme for water and sanitation in Central America: Facing new challenges and opportunities*, 96/11.*

Valdelin, J. *et al.*, Rural *Village Water Supply Programme – Botswana*, 96/10.*

Vadelin, J. *et al.*, *Sida Support to the Education Sector in Ethiopia*, 1992–1995, 96/27.*

Vance, I. and Vargas Cullell, J., *External Evalution of the Local Development Programme (PRODEL) in Nicaragua* (not published).

Whitmore, T. C., *Swedish–Malaysian Research Cooperation on Tropical Rain Forest Management*,96/18.*

Åhlund, C., *Democratic Development and Human Rights in Ethiopia*, 96/5.

Annex 6

Standard format for quality assessments

Sida Evaluations Data Worksheet

A. General Information

Report title	Report number Sida Evaluation

Authors	Professional background	Consulting firm/Institution

Sida Department and section responsible for evaluation

Other organizations responsible/funding the evaluation

Total cost of evaluation		Date of terms of reference	Date of final evaluation report	Has the project/programme been evaluated earlier?
Sida's share SEK	Total SEK			yes ☐ year(s) ____ no ☐ ____

B. The Project/Programme

Name of project/programme evaluated

Year of project start		Time period evaluated		Type of financing			
Total cost of project		Cost for period evaluated		Sida's share		Total	
Sida's share	Total	Sida's share	Total	grants	credits	grants	credits
MSEK	MSEK	MSEK	MSEK	%	%	%	%

Type of support

- Project support ☐
- Sector support ☐
- Programme aid ☐

Country/Region

____

____

not applicable ☐

Sector

- Social sector ☐
- Infrastructure ☐
- Economic sectors ☐
- Public administration ☐
- Disaster relief ☐

Please, specify :

____

____

____

____

____

Channel:		Implementing organisation		Recipient organisation	
Bilateral	☐	Public	☐	Public	☐
Multilateral	☐	Private	☐	Private	☐
NGO	☐	Mixed	☐	Mixed	☐
		NGO	☐	NGO	☐
		Research	☐	Research	☐

C. Type and Timing of the Evaluation

Timing		Type of evaluation (one alternative)			
Mid-term	☐	Project evaluation	☐	Programme aid evaluation	☐
End of project, completion	☐	Programme evaluation	☐	Thematic evaluation	☐
Ex-post	☐	Sector evaluation	☐	Organizational evaluation	☐
Not applicable	☐	Country evaluation	☐	Other	☐

D. The Evaluation Team

Was Sida personnel included in the team	yes	☐	no	☐
Was somebody from the recipient country/region included in the team	yes	☐	no	☐
Did the team consist of persons with specific sector competence	yes	☐	no	☐
Total number of evaluation team members/ thereof women		☐	/	☐
Total number of man weeks used for the evaluation/ thereof field work		☐	/	☐

E. The Evaluation Report

Report language	-------------------			
Executive summary included	yes	☐	no	☐
Summary language	-------------------			
Terms of reference annexed	yes	☐	no	☐
The team's workplan annexed	yes	☐	no	☐
Recommendations to be pursued by donor included	yes	☐	no	☐
Recommendations to be pursued by recipient included	yes	☐	no	☐
Lessons learned included	yes	☐	no	☐

F. Other Issues Addressed in the Evaluation

Sustainability issues	yes	☐	no	☐
Cost-effectiveness issues	yes	☐	no	☐
Gender equality aspects	yes	☐	no	☐
Environmental aspects	yes	☐	no	☐
Democracy and Human Rights aspects	yes	☐	no	☐
Poverty aspects	yes	☐	no	☐

Assessed by	Date

G. Presentation of Evaluation Methodology and Assumptions

Methodology issues are addressed	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Hypothesis and assumptions of the evaluation team are made explicit	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Discussions of validity of results	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Discussions of reliability of results	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Formal techniques used (e g CBA, ERR etc)	☐ Yes ☐ No

H. Methodology for Data Collection

(assessment of to what extent the various methods for data collection listed below have been used in relation to the evaluation task)

Agency documents used	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Use of other documents	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Agency personnel interviewed	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Project staff interviewed	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Beneficiary interviewed	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Other donor interviewed	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

Direct observation of project work			☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Participatory evaluation methods used			☐ Yes ☐ No ☐ Not addressed
Presentation of raw data			☐ Yes ☐ No
Interview protocols or guides used?	☐ Yes ☐ No ☐ Not addressed	Interview protocols or guides annexed?	☐ Yes ☐ No
Questionnaire used?	☐ Yes ☐ No ☐ Not addressed	Questionnaire annexed?	☐ Yes ☐ No

I. Analysis of Main Project Objectives

(assessment of to what extent the main determinants listed below have been discussed in the report)

Development objective	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Immediate objective achieved	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Key project assumptions examined	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

J. Analysis made in the Report

(assessment of to what extent the different types of analysis listed below have been carried out in the evaluation)

Country's need for the project	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Comparisons with/without the project	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Project expenditures analysis (costs)	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

Project benefit analysis	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Assessment of effectiveness	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Assessment of efficiency	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Discussion of the aid agency's performance	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Is there a discussion of lessons learned?	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

K. Sustainability of Project

(assessment of to what extent the following aspects of sustainability have been addressed in the evaluation)

Sustainability of project/programme after agency withdrawal	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Alternative financing of project/programme after agency withdrawal	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Degree of financing of project/programme from earned revenues	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Commitment of host country/local managment for sustainability	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

L. Cross-cutting Issues

(assessment of the treatment of additional issues ofgeneral character, as gender, environment etc)

Gender issues	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

Environment	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Human rights	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Democracy	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed
Market economy	☐ Exemplary treatment ☐ Addressed adequately ☐ Addressed minimally ☐ Not addressed

M. Concluding Overall Opinion about the Evaluation

	☐ Excellent ☐ Good ☐ Adequate ☐ Inadequate

Assessed by	Date for the assessment

Sida Studies in Evaluation - 1995/96 - 1997

96/1	Evaluation and Participation - some lessons. Anders Rudqvist, Prudence Woodford-Berger Department for Evaluations and Internal Audit
96/2	Granskning av resultatanalyserna i Sidas landstrategiarbete. Göran Schill Department for Evaluations and Internal Audit
96/3	Developmental Relief? An Issues Paper and an Annotated Bibliography on Linking Relief and Development. Claes Lindahl Department for Evaluations and Internal Audit
96/4	The Environment and Sida's Evaluations. Tom Alberts, Jessica Andersson Department for Evaluations and Internal Audit

Sida

SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY
S-105 25 Stockholm, Sweden
Tel. +46 (0)8-698 50 00. Fax +46 (0)8-20 88 64
Telegram sida stockholm Telex 11450 sida sthlm. Postgiro. 1 56 34–9
Homepage: http //www sida.se