

**T**

**owards an evaluation  
package. Review of 'What  
relief for the Horn?'**

Sophie Pinwill

**99-11**



Asia Pacific Press at the  
AUSTRALIAN NATIONAL UNIVERSITY

<http://ncdsnet.anu.edu.au>

© Asia Pacific Press 1999

This work is copyright. Apart from those uses which may be permitted under the Copyright Act 1968 as amended, no part may be reproduced by any process without written permission from the publisher.

The contribution of the Australian Agency for International Development (AusAID) to this series is gratefully acknowledged. The opinions contained in the series are those of the authors, and not necessarily of the National Centre for Development Studies or of AusAID.

ISSN 1443-4717

ISBN 0 7315 3625 8

**Sophie Pinwill** is a former student of the Graduate Studies in Development Administration program of the National Centre for Development Studies, The Australian National University. This paper was submitted at the close of the professional short course/optional Masters Degree module, *Complex Emergencies: humanitarian aid evaluation*, which was held by the National Centre for Development Studies.

## Abbreviations

SIDA Swedish International Development Authority

TOR terms of reference

## Introduction

Anne Marie Fallenius asserts

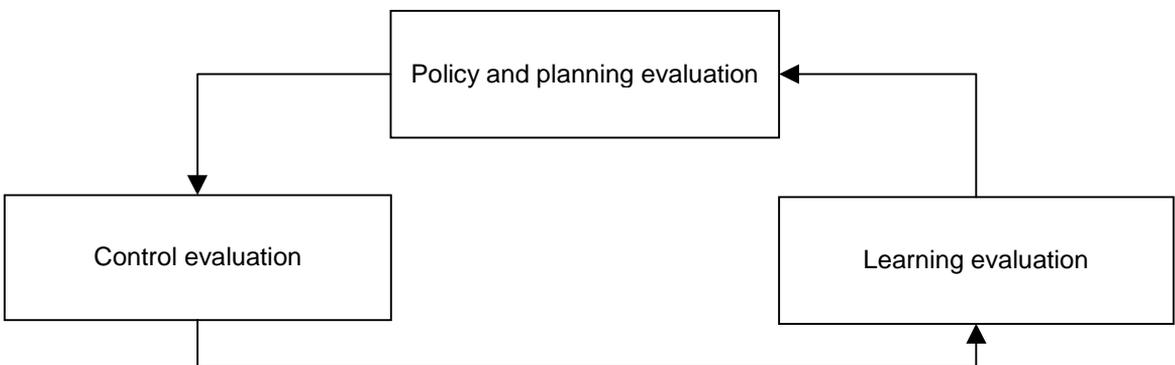
It might be useful to distinguish between two functions of evaluation.

1. Evaluation as an instrument of control (evaluation as audit)—to what extent have the goals of the intervention been achieved? was the intervention cost effective? can the aid resources spent be justified?
2. Evaluation as learning for future operations both for the specific activity and for the donor organisation (and recipients) in general(1996).

It would seem from the nature and characteristics of modern emergency relief operations—complexity, longevity, multi-dimensional in relation to causality and response—that a third function of evaluation needs to be added. This is a policy and planning review function. Complex emergencies, by definition, can not be planned for. This paper, however, will seek to show that with a policy and planning based evaluation within an evaluation package, the organisational context of relief aid can be explored without being hampered by other areas of investigation.

The evaluation conducted by a team led by Raymond Apthorpe in the Horn of Africa is, to a large extent, an example of a policy and planning evaluation. Although the reviewers acknowledged this from the outset, no formal recognition of this refined function was incorporated in the terms of reference (TOR), methodology, team selection or time frame. This meant that the Horn evaluation was, in effect, hampered by the complications of the other two types of evaluation. There are a number of methodological and other differences between a policy and planning evaluation and the kinds of evaluations with other functions, which will be outlined below.

## The evaluation package



This is a circular relationship based on the need for interconnectedness. The policy and planning evaluation should begin and end the ongoing process. Some potential characteristics of the circular relationship include the following process

1. Policy and Planning evaluation—setting the scene for a control evaluation, checking to see if systems are in place to be able to monitor and process data, check that field reports are done and there is administrative support for a control evaluation to take place.
2. Based on the findings of this part of the evaluation package, the terms of reference for a control and learning evaluation are designed, a time frame for both investigations is proposed, team composition decided and methodologies appropriate to the kinds of tasks required are planned.
3. The control evaluation would start when all of the available data and monitoring information had been collected and would then be able to assess cost–effectiveness, logistics, and so forth.
4. The results of the control evaluation could then serve as part of the guide for the learning evaluation to find out how improvements could be made to emergency operations.
5. The learning evaluation could then go on to explore such things as implementation strategies, impact, channels, and preparation.
6. All this information would then go back to the policy and planning review team and a follow–up analysis with recommendations would be made.

It is perhaps more useful to define the SIDA Horn evaluation in terms of what was achieved, rather than the limitations of the study. In order to do this, we must contemplate this third evaluation function and the benefits of thinking about evaluation as a package. Defining the limits and focus of the terms of reference for a policy and planning evaluation needs an alternative approach. This was not fully acknowledged by the Horn review team, and certain criticisms can perhaps be made as a result.

This paper will not argue the benefits of one type of evaluation over another, but seek instead to prove that there is a place in humanitarian and development assistance for all of them and that the relationship between the three approaches is complementary. This paper will further argue that the evaluation ‘package’ is dependent on the findings of all three components and is incomplete without them.

### **The package—why do we need a third category?**

In order to avoid some of the regular limitations and traps of traditional evaluation strategies, a design phase of evaluation is needed. Some of these limitations are outlined in the literature and include the broadness and lack of focus in the terms of reference of many evaluations, problems with team selection, lack of methodology and unrealistic time

frames. In his paper, 'The management of disaster relief evaluations', Claes Lindahl describes some of the problems associated with the evaluation conducted in Cambodia (Lindahl 1998). Many of his observations could be applied to evaluations generally, potentially helping to avoid some of these problems.

### **Terms of reference**

Lindahl describes the 'mutual disregard of the ToR by the agency commissioning the study and by the team carry (sic) out the evaluation...', which results from the unrealistic breadth of tasks and aims included in most TORs. It is impossible to cover the amount of ground required if one takes the standard TOR at face value (Lindahl 1998). At best, attention to all interest areas will be cursory and incomplete, at worst, areas will be ignored entirely. What commonly occurs is that TORs are refined to take account of their unrealistic scope, and, as Lindahl points out, this refinement is dictated by the agendas and interests of the evaluation team (admittedly, usually in consultation with the commissioning body). The question remains—how, then, are the gaps filled? Another crucial consideration, particularly in relation to SIDA and other funders who have a long term commitment to assisting regions plagued by a series of complex emergencies, is the addition of an assessment of future directional planning for the organisation. Why is this 'tacked on' to an already cumbersome and multi-dimensional terms of reference? This agenda should surely be a discrete and substantial area of investigation.

A policy and planning evaluation seen separately to a control or learning evaluation would have the task of designing a TOR for itself, and of providing a guide for the other two types of evaluation. The TOR for the policy and planning evaluation would include areas such as

- funder's commitment to and interests in the region
- relationships and cooperation with other funders,
- types of conflicts/contexts encountered
- policy directions in relation to target groups (such as women and children and demobilised soldiers)
- relationships with host governments
- aims of engagement (emergency aid, rehabilitation and development)
- relationships between these aims

*On the basis of the results*, the TOR would then assist in the design of the TOR for the other two types of evaluations. As Fallenius notes

[T]he composition of the evaluation teams, their backgrounds and preferences clearly have strong bearing on what evaluations will produce, to some extent (and possibly even to a high degree) independently of the ToR (Fallenius 1996:241).

This observation is very important and perhaps even more valid if the TOR is essentially unworkable. Problems with teams working independently of the TOR could be solved if the TOR were more achievable and less amorphous.

### **Team selection**

There seem to be some major problems associated with team selection in relation to both size and composition. In many cases, the selection criteria for the amorphous 'evaluate everything' approach appear enigmatic at best, and arbitrary at worst. Lindahl notes in the Cambodia evaluation that problems, or at least potential problems, arose from composition of the team, which was described as possessing 'a quite good mixture of competences' (1998:6), notably in terms of regional knowledge and experience. In spite of their competencies, Lindahl pointed out that the team members' performance suffered from a lack of methodology and that (more importantly), their extensive field experience in the region actually adversely affected their ability to conduct an objective evaluation. The team member's vested interests,

played themselves out in two different directions in the Cambodia evaluation in my opinion: 1) an apologetic view of the behaviour of the aid system in Cambodia in general and of the UN specifically, and some preconceived ideas of the value of various SIDA supported programmes, more based on personal affinities than facts; and 2) a tendency to look at programmes in which SIDA might continue in the future in an inherently positive way. In respect of the latter, the opportunities for future involvement by the team members concerned cannot be ruled out as a source of bias in the assessment (Lindahl 1998:7).

It would appear that regional experience and personal history in the area under scrutiny can lead to a conflict of interest and that, while local knowledge is important, precautions must be taken to ensure optimum objectivity. A policy and planning evaluation could provide some solutions to the vexing problems of team selection. Such an evaluation could help design selection criteria and define the types of expertise needed in all three types of evaluations, and, at the same time, look at ways to minimise bias and 'stakeholder syndrome' (Lindahl 1998:10).

### **Lack of methodology**

The main problem in the methodologies of the three SIDA evaluations, as discussed by Fallenius and Lindahl, was the lack of any methodological guidelines. It could be argued that the reason for this is that no single methodological approach is appropriate for the range of areas needing investigation, and that the evaluation teams were, in fact, correct in not assuming any one methodology and thereby potentially excluding certain levels of analysis. A policy and planning evaluation would have the luxury of considering the whole picture, while not having to answer all the questions. In this way, it could explore the specific terms of reference for the control and learning evaluations within the policy and organisational context, and design appropriate and separate methodologies to pursue a comprehensive evaluation.

## **Time frames**

Inevitably, evaluations have a habit of prefacing the study report by making 'lack of time available' an excuse (whether justified or not) for failing to look at certain issues in the TOR. While this indeed may be the case, it may be a misleading pretext. It is argued in this paper that evaluations as we know them suffer more from structural problems than lack of time. Addressing all the elements of a traditional TOR tends to fragment the evaluation, a problem not easily solved even with more available time. How are time frames for evaluations worked out? Is it based on a scientific or mathematical equation—number of people divided by number of countries, for example? Or is it resource driven, for example, 'we have X much money for this evaluation, see how far can you get'? How is the number of days in the field established, as opposed to days of preparation, or days spent in head office, or time writing up reports? A policy and planning evaluation can explore these issues of time and the timing of the three kinds of evaluation in the total package and devise a tool or standard set of guidelines for each type of evaluation.

## **The Horn evaluation**

### **Approach to the evaluation**

At the outset, a decision was made by the review team to do a policy review rather than a project evaluation. Given the time constraints and resources allocated to the task, it was impossible to perform a series of project evaluations (Apthorpe *et al.* 1994:81). It was clear upon reviewing the terms of reference, and also through preliminary consultations, that doing individual project evaluations would not be the best use of resources, or even appropriate, because some fundamental policy issues surfaced which needed to be investigated. At this point, the review team should have taken the opportunity to discard all elements of the proposed evaluation that did not fit into a policy and planning framework. However, in terms of some of the characteristics, the Horn evaluation falls into the same traps as other evaluations in relation to TOR—team selection, methodology, and time frames/timing. Before we begin a critique of these specific areas, there are some preliminary comments to be made.

### **Comments about the context of the evaluation**

From the outset it would appear that there was a certain lack of organisational commitment to this evaluation. The participation of the emergency relief section and the evaluation section was naturally guaranteed. Preparations, however, for the review team to meet partners in Sweden or hold discussions with other sections within SIDA were not arranged. This is perhaps indicative of the way evaluations are treated generally, not specifically within SIDA. An evaluation needs a strong and active advocate from within the organisation to 'oil the wheels' of the investigation. If there is not an active advocate, evaluation teams run the risk of being treated as if they are 'in the way' of the usual, admittedly busy and often hectic, office environment. Organisational commitment to an

evaluation on paper and in practice is obviously important in terms of carrying out the evaluation but it is perhaps even more important when it comes to following up recommendations and other post-evaluation exercises.

So the question must be asked—in spite of the request for, and commissioning of, the evaluation, is the organisation ready for an evaluation in terms of preparation and designation of sufficient time, resources and advocacy? How can the organisation better prepare for an evaluation in order to get the best and most accurate results? The value of a policy and planning evaluation in this context should not be understated. While looking at policy (although not exclusively) the Horn study picked up on the lack of organisational preparedness for a more comprehensive evaluation.

### **Terms of reference**

In the review document, the TOR of the Horn evaluation had already been edited and only an excerpt was provided (Apthorpe *et al.* 1994:60). Presumably the rest of the TOR which were not included in the excerpt included areas of investigation deemed impossible or inappropriate for this study. Even so, the remaining TOR framework falls prey to the usual problems of over-ambitiousness and fragmented directions of the scope and focus. The Horn evaluation, to its credit, did not attempt to analyse cost effectiveness or individual project performance, instead opting to focus on no fewer than thirteen separate areas of varying substance. If the Horn study had formally recognised the limitations of the overly expansive mandate dictated by the (revised) TOR, and instead employed the evaluation ‘package’ approach, the team would have more time to focus on policy and planning. As it was, the areas of interest on which the review subsequently focused were encumbered by the somewhat tokenistic regard paid to issues like ‘logistics’—an area that the team could not explore in any great depth, resulting in only ‘impressionistic’ knowledge.

### **Towards a policy and planning evaluation TOR**

A more fruitful and focused TOR for policy and planning evaluative purposes would include analysis of

- the history of, and intended future commitment of, the donor in the region, including National Policies with relation to trade, regional interests and nature of conflicts and emergencies.
- donor coordination
- cooperation with recipient countries
- purposes
- adjustment to changing circumstances
- rehabilitation and development aims
- policy level rationale for choice of target groups

## Team selection

It is unclear from the review report how the Horn evaluation team was selected. The report simply notes that team members had ‘considerable experience in the Horn and neighbouring countries (and in the case of the team leader elsewhere in the world as well) in the areas relevant to this review’s ToR’ (Apthorpe *et al.* 1994:82). There is no other information about which disciplines the team members came from, whether they were selected individually by the commissioning body, or indeed anything about the process of finding these people. This information is not usually included in evaluation documents except perhaps as a ‘team biography annex’, but the reason team selection is a subject of curiosity in this case is that it is not apparent exactly how each member contributed to the findings outlined in the report. More importantly, perhaps, could the team leader alone have conducted this type of policy evaluation? The section in the report dealing with organisation and method of the review (Apthorpe *et al.* 1994:81) states that all four members of the team did field investigations, and all members except the team leader did country reports, leaving Volume I (Thematic overview, conclusions and recommendations) for the team leader to write. How dependent was Volume I on the country reports? The earlier question remains (and only the members of the team can really judge)—could the same observations and conclusions have been reached by a single person team, doing the majority of the work at the donor’s headquarters, with assistance from a designated in-house evaluation advocate, and with more time? The benefits of ‘emancipatory reading’ (a phrase coined by Raymond Apthorpe—coincidentally the team leader of the Horn study) in respect of a policy and planning evaluation are numerous, if done in the context of the evaluation ‘package’. Reviewing policy and procedures is usually a matter of gaining access to organisational reports, consulting with staff, and having the capacity to analyse information from the field.

## Methodology

The methodology of the Horn evaluation was not discussed in the report. The travel schedules of the team members and the persons interviewed (but not the types of questions asked) were documented, as well as the references consulted. Nowhere, however, is the actual methodology of the evaluation outlined. This leads the reader to assume that the previous assertion—that there was indeed no methodology guiding the evaluation—is in fact the case.

If the Horn evaluation had not been ‘bogged down’ by the requirements in the TOR for control and learning types of evaluation, would it have had more chance of developing a relevant and specific methodological framework? How necessary was the field work part of the evaluation to the overall conclusions? Would a methodology review indicate that fieldwork was unnecessary, or that it could be done by one person as opposed to four?

## Time frame and timing

Before we discuss the time frame element in relation to the previous discussion in this paper, it seems that the actual timing of the evaluation was not particularly good. The fact that it was summer holidays in Sweden meant that many Swedish diplomats and officials were on leave, and consequently could not be canvassed for this evaluation. Also, the fact that the wet season in one of the countries precluded any field trips would seem to lead to the question of whether this was the best time to do the field studies. How much input did the evaluation team have in the actual timing of the event? Why was this particular time for the evaluation chosen? A better time could undoubtedly have been chosen, but the question remains: if indeed the Horn evaluation is a policy and planning one, is fieldwork even an integral part of the evaluation?

In terms of the time frame, the Horn review notes : ‘As is usual in the case of consultancy, recruitment was strongly driven by availability (at, in academic terms, a moment’s notice)’ (Apthorpe *et al.* 1994:82).

Why is this the case? Do commissioning organisations plan all their activities at a moment’s notice? Perhaps it is an indication of how little importance they afford evaluations that they would ‘rush it’ through. It would seem that better planning is necessary, as is a better understanding of exactly how much time is needed both to organise a team for evaluation and to actually conduct the evaluation. It is argued here that there is never enough time to conduct an evaluation, unless it is broken up into the three parts—policy and planning, control, and learning.

## Conclusion

This paper points out the limitations of traditional evaluation approaches. The all encompassing, multi-disciplinary, complex, time consuming, fragmented expectations of what an evaluation can achieve are unnecessary if the tasks are grouped into three natural areas. Much can be gained in terms of efficient use of time and resources by using the evaluation package approach. If the SIDA Horn review had aimed at doing a policy and planning review before the TOR were even drawn up, it would have had more time to focus on that area. The review would have also had input on the other two sections of the package, thus ensuring valuable follow-up was done. Without looking at the Horn evaluation in a package, the reader is left wondering what SIDA will do with the (valid and perceptive) recommendations.

## Bibliography

Apthorpe, R., Ketel, H., Salih, M., and Wood, A., 1994. *What Relief for the Horn? SIDA supported emergency operations in Ethiopia, Eritrea, Southern Sudan, Somalia and Djibouti*, SIDA, Stockholm.

- Fallenius, A.M., 1996 'Some methodological lessons learned from evaluation of disaster relief operations, Sweden', Paper produced for DAC Expert Group on Aid Evaluation Informal Seminar, Canberra, 6—8 February, January, 1996.
- Lindahl, C., 1998. 'The management of disaster relief evaluations: lessons from a SIDA evaluation of the complex emergency in Cambodia: some excerpts,' Management Perspectives International, Sweden, mimeo.
- United Nations High Commissioner for Refugees, 1998. 'Planning and organising useful evaluations', UNCHR Inspection and Evaluation Service, Geneva.