

# Evaluating Humanitarian Action

AN ALNAP GUIDANCE BOOKLET

September 2003 Draft 1

## Acknowledgements

This Handbook was produced following the guidance and with the support of the ALNAP Secretariat, and in particular Kate Robertson, John Borton, John Mitchell and Sera Orzel. It has greatly benefited from the constructive comments of a Review Panel consisting of:

Anne Bay Paludan, John Borton, Margie Buchanan-Smith, Andre Griekspoor, Alistair Hallam, Astri Suhrke, John Telford, Samuel Taddesse, and Peter Wiles.

**Tony Beck**

# Table of Contents

Acronyms .....	6
<b>1 Purpose and Use of the Guidance Booklet .....</b>	<b>7</b>
1.1 Background .....	7
1.2 Guidance Booklet users .....	7
1.3 Organisation of the Guidance Booklet .....	8
<b>2 Options for Evaluation of Humanitarian Action .....</b>	<b>9</b>
2.1 What is evaluation in the context of EHA? .....	9
2.2 Evaluation use .....	9
2.3 Lesson learning and accountability .....	11
2.4 Evaluating humanitarian protection .....	12
2.5 What is so special about EHA? .....	13
<b>3 Working with the DAC Criteria .....</b>	<b>14</b>
3.1 Effectiveness .....	16
3.2 Impact .....	21
3.3 Efficiency .....	25
3.4 Coverage .....	29
3.5 Relevance/appropriateness .....	34
3.6 Connectedness .....	38
3.7 Coherence .....	42
<b>4 References .....</b>	<b>47</b>
<b>Annex 1 American Evaluation Association Guiding Standards for Evaluation .....</b>	<b>49</b>
<b>Annex 2 Main types of evaluation of humanitarian action .....</b>	<b>52</b>
<b>Annex 3 Good practice in methods for the Evaluation of Humanitarian Action .....</b>	<b>53</b>



# Foreword

When we initiated the Joint Evaluation of Emergency Assistance to Rwanda in 1994 we had no manual, guidance or good practice to follow. The evaluation found that evaluations, for accountability or lesson learning, were few in numbers and, by many organisations, considered unnecessary: “We intend good, therefore we are good” and “you cannot put a price on human lives” (i.e. evaluate efficiency). Both arguments are totally invalid, and subsequent developments proved so.

The Rwanda Evaluation gave impetus to three initiatives (at least): (1) The SPHERE Project, setting standards for humanitarian action; (2) The HAP International, advocating accountability to beneficiaries. Both based on a philosophy of a rights based approach to humanitarian assistance, and (3) ALNAP. Together with the DAC Working Party on Aid Evaluation<sup>1</sup> ALNAP has championed evaluation as a tool for accountability and learning. The first product was the “RRN Good Practice Review: Evaluating Humanitarian Assistance Programmes in Complex Emergencies”, the second was the DAC “Guidance for Evaluating Humanitarian Assistance in Complex Emergencies”. Despite their commonality in titles, the two booklets aimed at different audiences, the first focussed on evaluators, the second on evaluation commissioners (managers).

Evaluation of Humanitarian Action has now become the norm rather than the exception. However, despite the validity and utility of the above mentioned two guide books the quality of evaluations of humanitarian action remains poor: the 2002 and 2003 ALNAP Annual Reviews reveal that while numbers have increased, quality leaves much to be desired.

This ALNAP Guidance: Evaluating Humanitarian Action builds on the best parts of its predecessors, and on the meta analysis of evaluations contained in the 2002 and 2003 Annual Reviews. It is a practical tool, pedagogically presented, and, if used properly, should improve the quality of future evaluations immensely.

**Niels Dabelstein**

Head of Evaluation Section DANIDA

<sup>1</sup> In 2003 renamed “The DAC Network on Development Evaluation”

# Acronyms

ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
DAC	Development Assistance Committee
EHA	Evaluation of Humanitarian Action
IDP	Internally Displaced People
LFA	Logical Framework Analysis
LRRD	Linking Relief, Rehabilitation and Development

# 1 Purpose and Use of the Guidance Booklet

## 1.1 Background

This Guidance Booklet has been developed after lengthy discussion in the evaluation community, and ALNAP in particular, about the need to strengthen evaluation of humanitarian action practice, and foster more effective use of the Development Assistance Committee (DAC) evaluation criteria.

The DAC evaluation criteria are currently at the heart of the evaluation of humanitarian action (EHA). A majority of evaluations include several or all of the criteria, and they are normally a central part of agency evaluation guides and manuals. However, several of the criteria are not well understood; and use of the criteria is often mechanistic, and excludes more creative evaluation processes (ALNAP 2002). The **objective of this Guidance Booklet** is to provide practical support as to how to use and apply the DAC criteria in the evaluation of humanitarian action, relative to particular contexts, types of intervention, and evaluation approaches.

This Guidance Booklet draws on a range of materials related to good practice in evaluation and EHA, including other guides, handbooks and manuals, which are referenced in the bibliography.<sup>2</sup> The intention was to build on, rather than replicate, what already exists. Development of the Booklet has also included a review of the widely used American Evaluation Association Program Evaluation Standards in terms of their relevance to EHA (see Annex 1). These standards are integrated into the main text where relevant; and the first standard – utility – is a cross-cutting theme.

## 1.2 Guidance Booklet users

Primary users are intended to be:

- evaluators of humanitarian action;
- agency staff involved in designing and managing EHA; and
- participants in training courses on EHA.

### Box 1

#### What are the DAC criteria?

*In 1991 the Development Assistance Committee (DAC) of the OECD set out broad principles for the evaluation process for DAC members. These principles were refined into five criteria that have been widely used in the evaluation of development initiatives – efficiency, effectiveness, impact, sustainability and relevance (OECD-DAC 2000). Subsequently the criteria were adapted for evaluation of complex emergencies (OECD-DAC 1999), and it is this document that is taken as a reference point here, focusing on seven criteria:*

■ Effectiveness ■ impact ■ efficiency ■ relevance/  
appropriateness ■ coverage ■ connectedness/  
sustainability ■ and coherence

*The DAC criteria aim to be a comprehensive set of measures against which interventions can be evaluated. They are complementary and can be used to evaluate all intervention elements related to results.*

<sup>2</sup> It also draws on a questionnaire completed by 25 ALNAP Full Members, and the work of the author on the ALNAP Annual Review.

Other intended users include developers of policy, programme design, and programme implementation. These groups have expressed a need for better understanding of the criteria against which interventions are being assessed. Their interests are addressed mainly through discussion of links between use of the Logical Framework Analysis (LFA) and the DAC criteria, in Section 3.

The DAC criteria, originally developed mainly for project evaluation, will not be equally relevant for all evaluations. Evaluations which have a policy or institutional focus may be less likely to use the DAC criteria. Annex 2 provides a typology of the main types/genres of evaluation currently undertaken, and the likely relevance of the DAC criteria in each case.

## 1.3 Organisation of the Guidance Booklet

Section 2 of the Booklet covers some of the main themes and issues in EHA over the last few years, with a particular focus on debates concerning lesson learning and accountability, and evaluation use, in order to contextualise the discussion of the DAC criteria in Section 3. Section 3 includes a definition of the individual criteria, an explication of the definition, issues to consider, key messages, and two good practice examples which are intended to provide pointers for those conducting evaluations. In addition to Annexes 1 and 2 already mentioned above, Annex 3 provides a brief note on good practice in methods in EHA.



# 2 Options for Evaluation of Humanitarian Action

## 2.1 What is evaluation in the context of EHA?

The definition used in this Booklet is the ALNAP definition, and given in Box 2.<sup>3</sup> However, there are many different types of evaluation (eg. internal, external, real time, ex-post) currently employed, and thus many different definitions of evaluation. Definitions related to humanitarian action tend to stress that evaluations are objective or impartial exercises intended to promote accountability and lesson learning. This Guidance Booklet is intended to support most types of evaluation.

## 2.2 Evaluation use

The Guidance Booklet aims to support an increased focus on evaluation use. Lack of use of evaluation findings and recommendations has been a concern for evaluation managers for some time. Evaluation planning therefore needs to pay greater attention to the final use and users of evaluation results, and plan accordingly. For example, if the primary audience is an Executive Board, the evaluation is likely to employ more formal evaluation approaches with a

focus on accountability; whereas if the main audience is primary stakeholders, a more participatory approach would be appropriate. Some practitioners have argued that unless evaluations are participatory and focus on process – that is engaging the evaluation audience from an early stage – then it is unlikely their findings will be used (Patton 1997).

A survey of 282 evaluators and evaluation managers (Preskill and Caracelli 1997) found that the most important strategies for facilitating use are:

- planning for use at the beginning of an evaluation;
- identifying and prioritising intended users and uses of the evaluation;
- designing the evaluation within resource limitations;

<sup>3</sup> In determining its definition of evaluation, there was considerable discussion among ALNAP Full Members concerning the potential for evaluations to be 'impartial'. Discussion of the impartiality of evaluation goes to the heart of one of the most important current debates on evaluation practice – that is how far evaluators can or should separate their own perspective from the evaluation process. This Handbook cannot cover this topic in detail, but interested readers can go to Lackey (1997) and Shulha and Cousins (1997).

### Box 2

#### What is the evaluation of humanitarian action? The ALNAP definition

A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability.

*It has the following characteristics:*

- *It is commissioned by or in co-operation with the organisation(s) whose performance is being evaluated.*
- *It is undertaken either by a team of non-employees (external) or by a mixed team of non-employees (external) and employees (internal) from the commissioning organisation and/or the organisation being evaluated.*
- *It assesses policy and/or practice against recognised criteria (eg. the DAC criteria).*
- *It articulates findings, draws conclusions and makes recommendations.*

- involving stakeholders in the evaluation process;
- communicating findings to stakeholders as the evaluation progresses; and
- developing a communication and reporting plan.

There are several questions to consider related to evaluation use and the DAC criteria:

- How much and what kinds of information do potential users need? Does there need to be equal focus on each of the criteria, or will some information be more useful than other information?
- When will the information be useful? For example, information on effectiveness and efficiency may be more useful in ongoing interventions.
- Can discussions about the evaluation terms of reference, including the DAC criteria, be used as an opportunity for evaluators to raise the issue of evaluation use?

Box 2.1 provides an example of an agency that has reportedly overcome many of the constraints to evaluation use.

## KEY MESSAGES

- ▶ It is the responsibility of evaluation offices and evaluators to ensure that evaluation findings are used.
- ▶ Commissioning agencies and evaluators need to start thinking about evaluation use from the planning stages of the evaluation.
- ▶ The evaluation process is often as important in promoting evaluation use as the end product.

### Box 2.1

#### Good practice example of promoting evaluation use: the case of Danida

*During the final phases of a Danida evaluation, evaluators are invited to present their key findings. During this presentation, stakeholders are invited to reflect on recommendations.*

*Danida has developed a form to facilitate the follow-up to evaluations in general, which is also applied to evaluations of humanitarian interventions. The form has three columns: recommendations; actions to be undertaken; and the status of this action.*

*When the final draft of the evaluation report is submitted to the evaluation department, this department transfers the recommendations in the relevant column of the form.*

*It then contacts departments concerned to formulate action to be taken for the follow-up to the recommendation. Recommendations and follow-up action are discussed in a Management Meeting, which is usually chaired by the State Secretary. During this meeting the recommendations and the follow-up action indicated are discussed, and adjusted or reformulated if this is required.*

*When agreement is reached, a decision is taken and authorized by the chairperson to execute the follow-up action as indicated. It thereby becomes an instruction to the concerned departments. Six months or more after the follow-up action has been decided, the evaluation unit approaches relevant departments to discuss progress in follow-up to the recommendations.*

*Source: van de Putte (2001)*

Table 2.1 Lesson-learning and accountability in EHA<sup>4</sup>

Evaluation approach	Lesson learning	Accountability
Examples of evaluation types	Self-evaluation, real time evaluation, process documentation, empowerment evaluation	Objective-based studies, impact assessment, ex post evaluation
Main current use	Intra-organisational learning	Reporting to funders, boards and the public
Main focus	Establishing why results were or were not achieved	Establishing whether results were achieved or not
Main methods used	Qualitative and participatory, e.g. Participatory Rural Appraisal	Quantitative e.g. collation of nutritional data
Use of DAC criteria	Less likely to use the DAC criteria	More likely to use the DAC criteria
Main problems	Information generated may not be considered credible because of lack of 'objectivity'	Tends to be non-participatory and may be viewed as a threat by those being evaluated

## 2.3 Lesson learning and accountability

As most definitions (e.g. IFRC 2002; UNICEF 2001; ECHO 1999) of evaluation stress the need for concurrent lesson learning and accountability, it is important to consider the implications of this for use of the DAC criteria. OECD/DAC (1999: 17) neatly summarises the implications of use of the two purposes:

A critical question to be considered at the outset is whether the evaluation is going to emphasise lesson-learning or accountability, or a mix of the two. If lesson-learning is emphasised then it opens up the possibility for the extensive use of participatory methods. If accountability is emphasized then it implies structuring the evaluation so that its findings are independent and respected.

In fact, proponents of these two purposes have been debating their relative worth in evaluation circles for some time (Patton 1997). Table 2.1 summarises the main elements and implications of each purpose.

Emphasis on one or the other purpose clearly has significant implications for selection of evaluation methods, including how and whether to use the DAC criteria. Evaluation genres, approaches and methods that better serve one or other of the purposes each have strengths and weaknesses, but a combination can provide an integrated focus on both process and results and serve both. This is why many working in the evaluation field, and in the social sciences in general, have been promoting a combined use of qualitative (read: lesson-learning) and

<sup>4</sup> Real time evaluation has been defined as a: 'timely, rapid and interactive peer review of a fast evolving humanitarian operation (usually an emergency) undertaken at an early phase. Its broad objective is to gauge the effectiveness and impact of a given ... response, and to ensure that its findings are used as an immediate catalyst for organizational and operational change.' (UNHCR: 2000, p. 1). Empowerment evaluation has been defined as: 'the use of evaluation concepts, techniques, and findings to foster improvement and self-determination. Empowerment evaluation has an unambiguous value orientation – it is designed to help people help themselves and improve their programmes using a form of self-evaluation and reflection.' (Fetterman 2002: 89). Process monitoring 'involves continuous information gathering over a period of programme work ... it is concerned with the dynamics of development processes, that means with different perceptions of relationships, transaction, decision making, or conflicts and their resolutions.' (Mosse et al 1998: 10) For a detailed discussion of models for evaluation, see Stufflebeam (1999).

quantitative (read: accountability) methods for some time (Kanbur 2001; Bamberger 2000). If used creatively, and as illustrated in Section 3, the DAC criteria can play a central role in this combined approach, and this Guidance Booklet intends to promote both accountability and lesson learning through use of the criteria.

---

## KEY MESSAGES

---

- ▶ When planning an evaluation, **evaluation managers and evaluators need to think carefully about how to combine lesson learning and accountability.**
- ▶ Are the necessary resources, including a mix of team skills, likely to be available?
- ▶ What is the main purpose of the evaluation, and will a focus on lesson learning and/or accountability mean that the evaluation results are more likely to be used?

## 2.4 Evaluating humanitarian protection

Humanitarian protection has emerged as a key issue in both the humanitarian response and EHA over the last few years. However, evaluation of humanitarian protection remains limited. For example, the ALNAP *Annual Review 2003* found that of 39 evaluation reports assessed, only two reports included adequate evaluation of protection.

ALNAP (2003: 13) defines humanitarian protection as: ‘the challenge of making states and individuals take up their humanitarian responsibilities to protect people in war and filling-in for them as much as possible when they do not.’ This definition goes beyond the approach taken by some agencies which focuses on improved programming so as to lessen risk; rather it suggests that humanitarian protection involves proactively establishing accountability vis a vis international human rights standards. In particular, this involves accountability to the affected population.

The monitoring of human rights violations is fairly well established, but as noted, the evaluation of protection remains in its infancy. This Handbook has attempted to be sensitive to ongoing debates as to the meaning of humanitarian protection, and whether evaluation of humanitarian protection can be mainstreamed in the DAC criteria. Issues related to evaluation of humanitarian protection are raised in Section 3 as relevant.

---

## KEY MESSAGE

---

- ▶ Evaluation managers and evaluators should always consider including protection as one area to be evaluated in EHA.

## 2.5 What is so special about EHA?

EHA operates as a sub-set of general evaluation, but has some distinct characteristics which need to be taken into account when planning to use the DAC criteria:

- Evaluations are often undertaken during periods of severe disruption, which in the case of complex emergencies can be prolonged. Access to key informants may be difficult;
- Conflicts polarise perspectives so that the same events are often subject to widely differing interpretations, diminishing the space for 'objective' evaluation;
- Data and information may be more difficult to come by; for example there is a high turnover of staff working in humanitarian action which may make it difficult for evaluators to interview key informants;
- Humanitarian action by its nature is often planned quickly, and objective statements and indicators may be missing from planning documents; and
- Evaluators often have to work through an interpreter.

---

### KEY MESSAGE

---

► **Lack of access to key information may make it difficult to employ all of the DAC criteria.** When completing the evaluation report, evaluators should make clear the constraints they faced and how these constraints affected the evaluation process and findings.

## 3 Working with the DAC Criteria

This Section outlines how the DAC criteria can be applied effectively in evaluation practice. The following is provided for each criteria:

- A definition;
- An explanation of the definition;
- Issues to consider in use of the criteria, including areas that can be problematic;
- Key messages; and
- Good practice examples, taken from a cross-section of countries, agencies, sectors, and natural disasters and complex emergencies. Further details on the good practice case studies, including in many cases full reports, can be found at the ALNAP Evaluative Reports Database - [www.alnap.org](http://www.alnap.org).

Table 3.1 at the end of the Section provides a summary of the main information on the DAC criteria for easy reference. The DAC criteria are designed to promote comprehensive evaluation of humanitarian action. For this reason, the criteria are complementary. For example, use of the criteria 'effectiveness' may show that objectives have been met, but this does not necessarily mean that the objectives are appropriate for the entire affected population, or were met efficiently. Similarly, an intervention by one agency can achieve good coverage, but may not be coordinated with other interventions. Using the DAC criteria in combination will ensure that the evaluation covers all areas of the intervention.

At the same time, evaluation managers will need to determine whether using the criteria in policy based or institutional evaluations is relevant. Also, as the criteria of impact, coherence and sustainability/connectedness cover wider issues related to social change, it may prove difficult to cover these criteria when evaluating single agency/single project interventions, such as a discrete water project or housing programme. The key to ensuring evaluation credibility here is to be transparent, that is to explain up front in the evaluation the constraints faced, and why a particular approach has been chosen

There will likely be some overlap in areas covered by criteria, such as impact and effectiveness, or effectiveness and coverage. It is up to the evaluator to decide under which heading to place information.

To evaluate interventions against the DAC criteria, it is preferable to have measurable objectives and baseline and results-oriented data. If these are not present, evaluators may have to construct them from available evidence, for example by interviews with key stakeholders, or oral history techniques.

There are a number of themes which should be considered in EHA which are not directly covered in the DAC criteria. These are areas important in understanding why humanitarian action worked or did not work, but which are not usually covered by the DAC criteria.

They are presented in checklist form in Box 3.1, and also highlighted for individual criteria when relevant. It will be useful to review this set of themes in the field and when writing up evaluation results. Themes included in Box 3.1 are illustrative rather than comprehensive, and

themes to be covered will always be evaluation specific. Not all evaluations need to include every theme, but if they are to be left out there should be a clear rationale for doing so. Analysis of themes during the evaluation will support analysis of why the intervention achieved particular results.

### Box 3.1 Checklist of themes to be covered in EHA

**The influence and understanding of local context** all intervention results are dependent, to varying degrees on national and local context, such as the security situation, availability of food in local markets, or the capacity of local institutions. When using the DAC criteria, evaluators should therefore consider the extent to which context was a determining factor.

**Human resources and management** a key factor in the success or failure of interventions, evaluation of human resources is currently one of the strengths of EHA (ALNAP 2002). Evaluators should pay attention to the level of experience/expertise of field staff; recruitment procedures; staff turn over; field/HQ relations and communication; the role of national staff; and training and learning practices.

**Participation of primary stakeholders** primary stakeholders need to be consulted about and to participate in all stages of interventions to ensure more ethical and effective interventions. But ALNAP (2002) found that primary stakeholders participated mostly in implementation, rather than in planning, design and decision making. This is a key area for evaluation, and further details can be found in the ALNAP Practitioner Handbook on *Participation by affected populations in humanitarian action* (ALNAP 2003). Evaluators and evaluation offices are also accountable to primary stakeholders, both in terms of consultation and in terms of ensuring that evaluation results are used to benefit primary stakeholders to the maximum possible.

**Coping strategies** the ability of primary stakeholders to manage emergency situations themselves is increasingly understood, but attention to this area is still limited in EHA. Evaluators should examine whether interventions have supported or hindered coping strategies, such as changes in nutritional practice, sale of assets, mutual support, or migration. Needs assessments also need to take into account livelihood and coping strategies. Gender equality: Many agencies have gender equality policies, which should be followed during response to crises. Evaluators should evaluate the extent to which interventions follow gender equality policies and promote gender equality. In relation to this, data in the evaluation report should be disaggregated by sex, where possible.

**The environment** Evaluations should assess whether interventions have supported environmental sustainability, or at the very least done no harm to the environment.



## 3.1 Effectiveness

<sup>5</sup> All definitions in this Section are taken from OECD/DAC (1999), unless otherwise noted. Many of these definitions draw on Minear (1994).

### Definition<sup>5</sup>

‘Effectiveness measures the extent to which an activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criteria of effectiveness is timeliness. Issues of resourcing and preparedness should also be addressed under this criteria.’

### Explanation of definition

Assessing effectiveness involves an analysis of the extent to which stated intervention objectives are met. For example, an objective statement might read: ‘Return the nutritional status of 50,000 female and male refugee camp dwellers to internationally accepted levels over a two month period, with an emphasis on the most vulnerable such as adolescent girls and the disabled.’

Assessing effectiveness in this case involves:

- examining the main reasons the intervention achieved or did not achieve particular objectives, including the process by which the change was brought about, for example: distributing food through community leaders and/or women; establishing food distribution committees which had gender balance in membership; or deploying and supporting effective agency staff;
- determining the change in nutritional levels of the target population over the stated period, using sex-disaggregated data where possible; and
- establishing that the change in nutritional levels was mainly caused by the intervention being evaluated. Preferably this would involve holding discussions with non-beneficiaries. At the very least there should be a plausible argument made concerning causality.

### Issues to consider

**Why were interventions effective, or not?** Knowing whether an intervention has met its objectives is half the story. The other half is knowing why this happened. In order to understand this, the checklist of cross-cutting themes will be a useful support – for example in asking questions related to who participated and why, and the influence of local context.

**Phrasing of results statements.** Many results statements are poorly phrased and not easily measurable. Objectives are often phrased as activities or inputs rather than as results, for example the number of houses to be built or amount food to be distributed, rather than the effect of the building or distribution. Evaluators should point out if this is the case; it is the role of the evaluator to attempt to evaluate the intervention against objectives, rather than activities.

Part of the assessment of objectives should involve analysing the process by which objectives were formulated, including who participated in this, and why. In particular, evaluators should ask whether and how primary stakeholders participated in the intervention’s design.

Where an LFA has been used it will normally be easier to assess effectiveness. Measuring effectiveness usually involves assessing the objective statements in the output column or row in the LFA.



**Use and benefit of resources.** Evaluations need to go where possible beyond assessing activities and begin to examine who uses and benefits from resources provided, with data disaggregated by sex, socio-economic grouping and ethnicity where possible. In this way a link can be made between the evaluation of effectiveness and impact. This may be a difficult area in ongoing operations, or where baseline or monitoring data is incomplete; even in these cases however primary stakeholder interviews can reveal important information on use and benefits.

**Timeliness.** A key element in the assessment of effectiveness is timeliness. The phasing of interventions is often crucial to success, and evaluations should consider whether interventions were carried out in a fashion that adequately supported the affected population at different phases of the crisis. Was timely arrival of support, goods and services achieved, according to the perceptions of different key stakeholders? The DFID/WFP good practice case below provides information on both the timing of the intervention as a whole, and distribution schedules for food aid.

---

## KEY MESSAGES

---

- ▶ Effectiveness measures agency objectives, which should be considered intermediate results at the output level
- ▶ When analyzing effectiveness, the evaluation should attempt to determine why the intervention has or has not achieved its objectives, and any lessons related to this for future interventions
- ▶ Understanding and analyzing the perspectives of primary stakeholders, and comparing these perspectives with those of other humanitarian actors such as agency staff, should be a key element in determining whether interventions have met their objectives

### 3.1.1 Coordination<sup>6</sup>

While not a 'formal' DAC criteria, EHA needs to pay close attention to coordination, which cuts across several criteria, but which is included here as a sub-set of effectiveness as coordination and effectiveness are closely related. Coordination can be defined as 'the systematic use of policy instruments to deliver humanitarian assistance in a cohesive and effective manner.' (Minear 1994).

Questions for evaluation of coordination include:

- Were plans for coordination in place and followed?
- Were there any incentives to coordinate, for example did donors promote UN coordination through funding arrangements? Or was there competition for funds?
- Was a lead agency appointed, and what was the result of this?
- Which parties were included and in what manner? Why?
- Who took the lead and how effective were they? Why?
- Were funds channeled in a coordinated fashion, or individually by donors to suit their own strategic aims?

<sup>6</sup>This Section draws on Reindrop and Wiles (2001) and Van Brabant (1999).

## Good practice in assessment of effectiveness 1

### DFID evaluation of support to WFP in Bangladesh

#### Background

In September 2000 about 2.7 million people were seriously affected by floods in six southwestern districts of Bangladesh. DFID supported WFP in providing a full ration of rice, pulses and oil to 260,000 beneficiaries during a first distribution and 420,000 during a second and third distribution. The DFID evaluation provides a comprehensive analysis of whether project objectives were met in relation to distribution of food aid, with particular reference to ration sizes, commodity mixes, and distribution schedules, the latter being one of the factors contributing to timeliness.

#### Choice of method to evaluate effectiveness

**The evaluation included both quantitative and qualitative methods** Quantitative data was collected in 2,644 randomly selected households in villages throughout the project zone. Qualitative data was collected during livelihood assessments in six representative villages on the livelihoods systems, status and prospects in flood-affected communities. A second smaller evaluation team was deployed about five weeks after the end of the first qualitative assessment to explore community perceptions and behaviours related to the food ration, including issues such as timeliness of distribution, desirability of commodities, and usage patterns. The quantitative and qualitative data sets were used in combination in the analysis.

#### The report includes most key elements for the evaluation of effectiveness

- ▶ examination of development of the intervention objectives, including analysis of the LFA;
- ▶ assessment of criteria used for selection of beneficiaries, including primary stakeholder views of these criteria - an area which can also be assessed under the DAC 'coverage' criteria;
- ▶ analysis of implementation mechanisms, including levels of community participation;
- ▶ targeting accuracy, disaggregated by sex and socio-economic grouping;
- ▶ examination of the resources provided – both the size of the ration and the commodity mix - which can also be assessed under the DAC 'relevance/appropriateness' criteria - including the reasons why they were provided;
- ▶ adequacy of distribution schedules; and
- ▶ the affected population's view of the intervention.

**Source** DFID (2001) *Emergency Food Aid to Flood-Affected People in South-western Bangladesh: Evaluation report*. (London: DFID)

## Good practice in assessment of effectiveness 2

### WFP evaluation of Food Assistance and Support for Repatriation of Iraqi and Afghan Refugees in Iran

#### Background

WFP has been active in Iran for 15 years at the time of the evaluation, providing food assistance to Afghan refugees since 1987 and to Iraqi refugees since 1988. This evaluation of a Protracted Relief and Recovery Operation (PRRO) was carried out over four weeks in early 2002 by two consultants and two WFP staff, assisted by two consultants. The evaluation team visited seven of the 29 refugee camps currently assisted by WFP. The main objective of the evaluation was to assess the efficiency, relevance, effectiveness, coherence and sustainability of WFP assistance under the PRRO in order to improve the implementation of the current operation and assist with planning the next phase.

#### Evaluation of effectiveness

The evaluation report first clearly sets out the four stated goals/objectives of WFP food assistance, and goes on to assess the extent to which outputs related to these goals have been met. For example, the first goal is to: 'Ensure the basic food needs for survival of the refugees', one corresponding outputs of which was: 'Efficient distribution of a balanced food basket equivalent to 1,900 kilocalories to a target of 84,000 beneficiary refugees.' The evaluation notes percentage of food aid delivered against assessed requirements for each of the main commodities in total, and then examines variations between camps. It goes on to examine why there were shortfalls, including the political environment and planning processes.

This clear delineation of intervention outputs and in-depth analysis of whether and why outputs have been achieved is at the heart of the evaluation of effectiveness. This should be linked where possible to the evaluation of impact; for example the WFP evaluation report goes on to examine whether the intervention has led to improved nutritional outcomes, going beyond the evaluation of effectiveness to questions of distribution and use of commodities.

**Source** WFP (2002) *Full Report of the Evaluation of IRAN PRRO 6126 – 'Food Assistance and Support for Repatriation of Iraqi and Afghan Refugees in Iran.'* (Rome:WFP)

- What were the main constraints and supports to coordination? How was good coordination achieved, and is it replicable for other situations.
- Did coordination lead to improved effectiveness and impact, for example an increase in humanitarian space? How did this come about?

#### Issues to consider

**The multiplicity of actors.** Assessment of coordination is made difficult because of a multiplicity of actors and the various potential routes of coordination among them (eg between donors, between donors and NGOs, between donors and the UN system etc). But numerous evaluations point to coordination as a problematic area which will pay dividends if improved (ALNAP 2002), and hence it should be included in evaluations and lessons learning exercises where possible.

**The role of the host government and other local institutions** are important to consider, for example whether objectives and activities were in line with host government priorities. Host governments are often bypassed which means that local capacity is not built, but many governments now have bodies in place to coordinate humanitarian action (Van Brabant 1999).

**Non-traditional partners.** A further issue concerns how coordination with non-traditional partners such as the military was organized. It is important to solicit stakeholder viewpoints on this type of coordination.

## 3.2 Impact

### Definition

‘Impact looks at the wider effects of the project – social, economic, technical, environmental – on individuals, gender and age-groups, communities, and institutions. Impacts can be immediate and long-range, intended and unintended, positive and negative, macro (sector) and micro (household).’

### Explanation of definition

Whereas assessment of effectiveness examines whether intervention outputs have been met and objectives achieved, assessment of impact usually examines the longer term consequences of achieving or not achieving those objectives and the issue of wider social change. For example, effectiveness would examine whether nutritional levels had improved, and impact would analyse what happened if nutritional levels did or did not improve, for example beneficiaries being able to undertake work. As such, assessment of impact often goes beyond intervention planning documents to consider the part the intervention plays in wider socio-economic and political goals – as can be seen in the good practice case from Rwanda. Because of its longer-term focus, evaluation of impact and of sustainability are often closely linked.

### Issues to consider

**Is evaluating impact relevant in all cases?** Because of this wider scope, assessment of impact may not be relevant for all evaluations, particularly those carried out during or immediately after the intervention. Changes in socio-economic and political processes may take many months or even years to become apparent. Also, assessment of impact may need a level of resources and specialised skills which have not often been deployed in evaluations of humanitarian action to date. Therefore, evaluation of impact should only be attempted where a longitudinal approach is being taken, or where there is data available to support longer-term analysis, the evaluation team includes specialists in socio-economic and political analysis, and the commissioning agency is willing to invest in a more detailed evaluation.

**How to deal with attribution?** The question of attribution may need special attention in the assessment of longer-term change. The further one moves from the time of the intervention, the more difficult it is to determine if the changes that have taken place are the result of the intervention or some other factor, such as other interventions, or socio-economic or political forces. For this reason ‘informal’ control groups of the affected population who have not received assistance should be interviewed.

**Phrasing of results statements, and the LFA.** Impact can be considered a higher order measure and thus relates to the goal and purpose columns in the LFA. As such it is more likely to be able to address longer-term goals such as support of human rights or gender equality, changes to political and social structures that may have led to the crisis, or support of livelihoods. Goal and purpose statements in the LFA are often vaguely worded with few quantitative targets stated. Evaluators may need to refer to agency policy to determine how the intervention fits with longer-term goals of the agency.

**Livelihoods.** Evaluation of impact should always consider support provided to livelihoods of primary stakeholders, in particular longer term adaptive strategies.

## Good practice in evaluation of impact 1

### Joint Evaluation of Emergency Assistance to Rwanda

As noted above, the evaluation of impact requires a broad focus on the consequences of an intervention. To date the largest and most comprehensive evaluation of humanitarian action, involving 52 consultants and researchers, the Joint Evaluation of Emergency Assistance to Rwanda (JEEAR) also sets standards for the joint assessment of impact of political action and its lack in complex emergencies.

JEEAR assesses impact mainly in terms of a lack of intervention in Rwanda by the international community, in spite of significant signs that forces in Rwanda were preparing the climate and structures for genocide and political assassinations. As such it employs a definition of humanitarian action that includes both political and socio-economic functions; this necessarily leads to an analysis of political structures which to a large extent determine humanitarian response and impact.

In the Rwanda case the lack of intervention is considered in two parts:

- ▶ an analysis of historical factors which explained the genocide; and
- ▶ a detailing of the immediate events leading up to the genocide.

The value of the joint evaluation is that it allowed an assessment that went beyond the confines of examination of single sector interventions to analysis of political economy. The political economy approach is then linked to the evaluation of the effectiveness of the humanitarian response.

This approach can be contrasted with two other crises: conflict and its aftermath in Kosovo, and the effects of Hurricane Mitch. In each of these cases decisions were made to carry out single agency and single sector evaluations, which missed to a large extent the political nature of the event and the response to it. In the Kosovo case this led to a lack of attention by evaluators to issues of protection (ALNAP 2001); and in the Central American case it led to lack of attention to how far humanitarian action supported the transformative agenda proposed in the Stockholm Declaration (ALNAP 2002).

JEEAR is unusual in its assessment of impact because it places a strong emphasis on why there was little interest in intervening in Rwanda - principally because of its lack of geo-political significance - rather than listing what events unfolded and their consequences. One of the lessons for evaluators of JEEAR is that evaluations of impact need to look not only at what interventions took place, but also what might have happened given other circumstances and different kinds of intervention. In other words, evaluations of impact may need to take into account the hypothetical as well as the actual.

**Source** JEEAR (1996) *The International Response to Conflict and Genocide: Lessons from the Rwanda Experience*. 5 volumes. Copenhagen: Steering Committee of JEEAR

## Good practice in evaluation of impact 2

### Evaluation of ECHO health, nutrition, water and sanitation in Sierra Leone

#### Background

This sectoral evaluation was part of a global evaluation of ECHO's Global Plan in Sierra Leone in 2000 and 2001. The Global Evaluation was carried out by a team of three consultants, with the health, nutrition and sanitation report being written by one of these consultants. The purpose of the evaluation was to both assess the suitability of ECHO's operations and to make recommendations for future programming on the basis of lessons learned. DAC criteria evaluated were relevance, efficiency, effectiveness, impact and sustainability. The evaluation was based on a standard methodology including: document review; interviews with key agency staff; and visits to 11 projects. Evaluation constraints were noted as: lack of reliable population data, the difficulty of constructing a narrative and baseline due to the high turnover of humanitarian agency personnel, and weaknesses in project monitoring and reporting which made the assessment of quantitative data difficult.

#### Evaluation of impact

Evaluation of the ECHO intervention is well grounded in analysis of political, social and economic context, as well as the sectors covered, and also notes the challenges of measuring impact in the humanitarian context. However, using available data the evaluation discusses the prevalence of malnutrition and the effects of immunization. Building on the discussion of these basic statistics, it goes on to consider the wider impact of ECHO's intervention on the health care system, and public services. It shows that even with humanitarian action's short time frame and lack of data, it is possible to examine impact.

In addition to examining quality of life indicators, the evaluation ties in discussion of context to some of the longer term impacts of ECHO funded work. It notes that ECHO projects have mitigated the effects of conflict by financing operations of a stabilising nature and improved the capacity of local communities to integrate IDPs and returnees. It also considers possible unintended consequences, including:

- ▶ The effect on the government budget, as financial support will leave the Ministry of Health Services with long-term expenditures that it cannot afford;
- ▶ Delivering aid in rebel areas has involved negotiating with rebel leaders, which has conferred on them a certain legitimacy which may reinforce existing power structures; and
- ▶ A dependency syndrome is being created with NGOs dominating the direction of health care and government counterparts at district level having limited involvement in priority-setting and planning.

Continued

## Good practice in evaluation of impact 2 continued

### Evaluation of ECHO health, nutrition, water and sanitation in Sierra Leone

Reasons for achieving impact are also discussed in some detail, for example the evaluation notes that (p14): 'Experience elsewhere has demonstrated that a comprehensive approach can influence the health status of a community and be more effective than establishing the infrastructure only. Over the past year ECHO-funded projects have gone beyond building wells, latrines and installation of hardware. A combination of approaches including hygiene promotion, safer disposal, hand washing and maintaining drinking water free of contamination has been adopted...'. The report also notes the importance of community involvement in the achievement of impact.

Source: ECHO (2001) *Sierra Leone, ECHO Global Plan 2000/Intervention Plan 2001. Health, Nutrition, Water and Sanitation – 2001*. Brussels: ECHO.

**Protection.** The evaluation should consider whether the intervention led to an increase in humanitarian space, greater protection for the affected population, and an advancement of human rights. Were there deliberate, structured interventions designed to improve protection, as well as inadvertent or semi-planned effects of other interventions?

---

## KEY MESSAGES

---

- ▶ A key element in the assessment of impact is explaining why events happened. That is, what were the processes involved that led to particular results, and whether positive results are replicable.
- ▶ Assessment of impact may be the most challenging aspect of carrying out an evaluation. When including impact as an area of enquiry in an evaluation terms of reference, evaluation managers need to consider whether the evaluation team has the relevant skills and resources to evaluate impact adequately.
- ▶ Assessment of impact can be conceptualized as a process of understanding what has changed at the individual, household and community levels, as well as the regional, national and international levels.



## 3.3 Efficiency

### Definition

**‘Efficiency** measures the outputs — qualitative and quantitative — in relation to the inputs. This generally requires comparing alternative approaches to achieving the same output, to see whether the most efficient process has been used. **Cost-effectiveness** is a broader concept than efficiency in that it looks beyond how inputs were converted into outputs, to whether different outputs could have been produced that would have had a greater impact in achieving the project purpose.’

### Explanation of definition

Efficiency usually measures how economically inputs (usually resources such as funds, expertise, and time) were converted to outputs. Assessment of efficiency tends to start with financial data, and should factor in the urgency of assessed needs of the affected population.

Determining whether the intervention was implemented in the most efficient way may involve comparison with alternatives – eg, providing piped water rather than trucking water in, supplying goods by road rather than air, and using food aid rations that are more culturally appropriate and therefore more likely to suit the needs of the affected population.

### Issues to consider

**Political priorities of governments and agencies may cause interventions to be inefficient.** For example, a host government may not want piped water provided to refugees if it does not want to encourage them to stay on its territory; or a donor may want to supply goods by air as this provides it with a higher profile in the media. Evaluators therefore need to take political factors into account.

**What was the source of inputs?** Part of the assessment of efficiency considers whether goods/inputs were purchased most efficiently in relation to source of input. One key question is whether inputs were locally purchased or imported. For example an evaluation of interventions in East Timor notes that supplies were procured in Geneva rather than more efficiently in Darwin. A related question is whether local tenders were sought.

**Financial areas to consider** are: total cost of the intervention broken down by sector; costs of inputs locally and internationally; transportation costs broken down by sector and type of transportation; staff costs, broken down by local and expatriate staff; and administration costs as a percentage of intervention costs. Evaluation of efficiency may require the inclusion on the evaluation team of a person with an economics or accounting background.

Efficiency is mainly covered in the input and output columns or rows of the LFA. In some LFAs financial information is included at the input level, and this will provide a direct lead into the evaluation of efficiency.

## Good practice in assessment of efficiency 1

### Evaluation of Disasters Emergency Committee Mozambique Flood Appeal Funds

#### Background

The Disasters Emergency Committee (DEC) evaluation of its intervention in Mozambique after the 2000 floods takes a close look at the humanitarian response undertaken by the responding DEC agencies, in order to report to the UK public on how and where their funds were used, and to identify good practice for future emergency operations. The method for the evaluation included extensive interviews, background research and field visits, and a detailed beneficiary survey.

#### Evaluation of efficiency

The chapter dedicated to efficiency contains many of the key elements necessary for evaluation. This included analysis of:

- ▶ use of military assets by the DEC agencies, assessed in terms of lack of collaborative use of helicopters to carry out needs assessment; the high costs of using western military forces for humanitarian relief, as compared to use of commercial facilities; and the comparative costs of the Royal Air Force, US military and South African National Defence Forces. The report notes that expensive military operations consumed large amounts of funding, which limited later donor funding of NGO projects;
- ▶ The effects on efficiency of an underdeveloped market for contracted services; for example, although use of national contractors enabled agencies to implement equipment-heavy works, such as road repairs, without having to make large capital investments, the contractors used by the DEC agencies often failed to meet their obligations in a timely manner;
- ▶ The efficiency of choice of response, i.e. intervening directly with operational programmes, working through local partners, or working through international network members. The evaluation found that staff composition was a more important factor determining efficiency than choice of response;
- ▶ Whether it was more efficient for agencies to build their response on existing capacity in-country or through international staff;
- ▶ Whether agencies with existing partners were more efficient than those without;
- ▶ How investment in preparedness led to a more efficient response;
- ▶ The efficiency of accounting systems;
- ▶ An attempt was made to compare input costs between the different agencies, for example of emergency kits, but this proved impossible given the different items provided and delivery channels used. Instead, the evaluation relied on cost implications of general practice followed, such as warehousing practices and transportation costs.

Continued

## Good practice in assessment of efficiency 1 continued

### Evaluation of Disasters Emergency Committee Mozambique Flood Appeal Funds

In addition the evaluation includes a breakdown of expenditure of funds by sectors, and for each of the DEC agencies by supplies and material, non-personnel and personnel, and agency management costs.

Source: DEC (2001) *Independent Evaluation of DEC Mozambique Floods Appeal Funds: March 2000-December 2000*. London: DEC.

## Good practice in the evaluation of efficiency 2

### UNHCR evaluation of the Dadaab firewood project in Kenya

#### Background

This is a thorough and thoughtful evaluation of a UNHCR project initiated in 1997 primarily to address issues of rape and violence against women and girls, with secondary objectives of environmental rehabilitation and reducing resource-based conflicts between refugees and local communities. Like many refugee situations, the Somali population in the camps supported by UNHCR are subject to protracted dislocation which had lasted for a decade at the time of the evaluation. Unusually this evaluation examines in considerable depth both institutional, socio-economic and environmental issues, linking these three areas together. The method involved document review, interviews with key staff, and beneficiary interviews.

#### Evaluation of efficiency

The evaluation includes a detailed discussion of the costs of the firewood project in relation to other possible uses of funds (pp99-110). Firstly the evaluation notes how it was hampered by lack of accurate financial data (p100): 'It has been exceedingly difficult for the consulting team to determine the actual costs of the firewood provision, and therefore to calculate costs per head, or costs per month, or total costs per ton of firewood provided ...' It is important that evaluations make such constraints clear. Then the evaluation goes on to examine key areas in relation to efficiency:

Continued

## Good practice in the evaluation of efficiency 2 continued

### UNHCR evaluation of the Dadaab firewood project in Kenya

- ▶ The cost of the firewood project compared to self-collection and existing commercial supply systems;
- ▶ Firewood supply in relation to other Dadaab refugee support spending;
- ▶ Firewood costs in relation to other gender-based violence reduction options, for example increased police patrols within and around the camps, and better training and sensitisation of police, improved fencing around the camps, identification of banditry prone areas, and improved lighting. Each of these options, including cost, are considered in detail; and
- ▶ Alternative ways of reducing firewood related trips, for example improved stoves, collective cooking and fire management.

As a result of this analysis the evaluation recommends alternative more cost-effective options which are also likely to reduce the extent of rape and sexual and gender-based violence.

Source: UNHCR (2001) *Evaluation of the Dadaab firewood project, Kenya*. Geneva: UNHCR.

---

## KEY MESSAGES

---

- ▶ Given that many humanitarian interventions involve the provision of large quantities of material inputs, analysis of efficiency is important in ensuring that resources have been used appropriately, and also has the potential to highlight more effective use of resources.
- ▶ One of the purposes for measuring efficiency is to ensure that all of the resources provided in an intervention have been used to the best effect. As with the other DAC criteria, a key question to ask is why the intervention reached its level of efficiency, for example in relation to planning, expertise of staff in logistics, or policies on purchasing.
- ▶ The rush to respond to a crisis for political reasons or the need for a high profile, and subsequent inadequate needs assessment, has often meant that resources are not provided in an efficient manner. It is the role of evaluators to highlight such poor practice, as well as to draw attention to any good practice and how this might be replicated.

## 3.4 Coverage

### Definition

‘The need to reach major population groups facing life-threatening risk wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agendas.’

### Explanation of definition

Evaluation of coverage involves determining who was supported by humanitarian action, and why. In determining why certain groups were covered or not, a central question is: ‘What were the main reasons that the intervention provided or failed to provide major population groups with assistance and protection, proportionate to their need?’

Assessment usually works at three levels:

The international level, determining if the resources provided in one emergency are adequate in comparison to other emergencies. Evaluations could spend more time considering this issue. A good example is DEC (2000) which examines this issue in relation to the Kosovo conflict;

- At the national or regional level, determining whether support was provided according to need of different areas, and why or why not; and
- At the local (e.g. village, slum, community and/or refugee camp) level, determining who received support and why; information at this level should be broken down by social categories such as socio-economic grouping, gender, age and ethnicity.

As the definition makes clear, evaluating whether protection needs have been met is a key element in the assessment of coverage. Even where protection issues do not form a major part of the intervention, evaluators should still assess whether protection issues should have been integrated into planning.<sup>7</sup>

Evaluators need to assess the extent of **inclusion bias**, that is inclusion of those in the groups receiving support who should not have been (disaggregated by sex, socio-economic grouping and ethnicity); as well as the extent of **exclusion bias**, that is exclusion of groups who should have been covered but were not (disaggregated by sex, socio-economic grouping and ethnicity).

The Principles of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Response Programmes are clear concerning who should receive support (see Box 3.2). DEC (2002) provides an innovative example of how the Code can be used as an evaluation tool.

Box 3.2

### Red Cross/Red Crescent Code of Conduct and coverage

*Aid is given regardless of the race, creed or nationality of the recipients and without adverse distinction of any kind. Aid priorities are calculated on the basis of need alone. Wherever possible, we will base the provision of relief aid upon a thorough assessment of the needs of the disaster victims and the local capacities already in place to meet those needs. Within the entirety of our programmes, we will reflect considerations of proportionality. Human suffering must be alleviated whenever it is found; life is as precious in one part of a country as another. Thus, our provision of aid will reflect the degree of suffering it seeks to alleviate.*

Efficiency

Coverage

<sup>7</sup> A recent paper on needs assessment and protection argues that protection needs should be taken as the fundamental framework for analysis in conflict situations but the current tools for current this approach may not be adequate (ODI 2003).

### Issues to consider

**Differing perspectives on what constitutes need.** What constitutes 'need', and therefore who is covered by humanitarian action, is often culturally determined. There are differing opinions as to whether the whole of an affected population, or the most vulnerable within that population, should be covered (ALNAP 2002). Evaluators need to be sensitive to this issue and determine if cultural practices in the intervention, often determined by donor governments, are appropriate from the perspective of primary stakeholders.

The evaluation report should present an estimate of the proportion of those in need covered, expressed as a percentage, rather than an absolute number. Reports in the past have tended to provide an absolute figure of those covered, which gives no sense of the total population in need.

**The situation of Internally Displaced People (IDPs)** may need special attention when evaluating coverage, as IDPs are not included in the international law that protects refugees, but may be among the most needy of the population. A good example of an evaluation examining the coverage of IDPs is Danida/UNHCR (2001).

**In the LFA**, the coverage area is usually incorporated in results statements and indicators relating to the numbers and types of the affected population targeted. Results statements should be clear concerning the numbers to be covered, as well as particular groups which are being targeted. Terms such as 'vulnerable groups' should be disaggregated.

---

## KEY MESSAGES

---

- ▶ Evaluating proportionality, that is evaluating whether aid had been provided according to need, is central to the evaluation of coverage.
- ▶ Coverage is often determined by political factors, and understanding why certain groups were covered requires an analysis of these factors, often including issues of protection and humanitarian space – as in the DANIDA good practice example.
- ▶ Equity questions are central to analysis of coverage. Coverage should consider equity through both geographical analysis and a breakdown of data by relevant socio-economic categories, e.g. gender, socio-economic grouping, ethnicity, age and disability.

## Good practice in assessment of coverage 1

### Evaluation of Danish Humanitarian Assistance to Sudan 1992–1998

#### Background

This evaluation covers Danish support to Sudan, one of the largest recipients of Danish humanitarian assistance, during the period 1992–1998. As one of several evaluations assessing Danish humanitarian assistance around the globe during the 1990s, this evaluation concentrated in particular on coverage.

#### Evaluation of coverage

The evaluation is considered good practice because it is marked by an in-depth analysis of the politics of coverage, including questions of humanitarian space, which an evaluation looking at a period of a number of years is well placed to assess. It usefully contrasts interventions which attempt to deal with short-term needs (e.g. basic health and nutrition) as opposed to longer-term structural problems such as the marginalisation and exploitation of displaced communities. The evaluation views the latter as a key factor in determining the extent of coverage. It also includes a number of key areas for assessment of coverage which other evaluations could follow:

- ▶ the overall context including the numbers in need and current provision of resources from external support;
- ▶ overall funds devoted to Sudan;
- ▶ standards of coverage in Sudan; according to the evaluation, these were revised downwards by agencies to lower levels than is usually considered the norm, thus decreasing the numbers supposedly 'in need';
- ▶ total levels of food aid supplied, including a detailed discussion of humanitarian space and how insecurity and political obstacles have limited the provision of aid;
- ▶ DANIDA's input into the increase of humanitarian space;
- ▶ Assessment of coverage disaggregated by ethnicity; and
- ▶ Assessment of coverage of internally displaced persons, including analysis of why this group is not being covered in some cases.

**Source** DANIDA (1999) *Evaluation of Danish Humanitarian Assistance 1992-1998: Volume 7, Sudan* Copenhagen: DANIDA

## Good practice in assessment of coverage 2

### WFP evaluation of food aid for relief and recovery in the Great Lakes

#### Background

The purpose of this report was to provide accountability to the WFP Executive Board, and to assess the usefulness of the Protracted Relief and Rehabilitation Operation (PRRO) as a resource mechanism and programming instrument in the Great Lakes and as an effective tool for supporting relief and recovery activities in the region. The evaluation team was made up of four expatriates, one WFP staff and three consultants, who visited the region for five weeks in early 2002. The Great Lakes PRRO at the time of the evaluation was WFP's largest, targeting 1.12 million people annually with a planned total budget of US\$437 million. Given the size of the programme and its regional nature, evaluation of coverage offered a particular challenge. This was heightened by the typical lack of data in such evaluations (p17): 'The evaluation mission found it very difficult and time consuming to obtain data on food distributions by component, on beneficiaries, or to measure achievements and progress against what had originally been approved.'

#### Evaluation of coverage

The terms of reference for the evaluation include a useful set of questions in relation to targeting and assessment. Targeting is covered in a comprehensive manner from geographical, political, community and intra-household perspectives. The evaluation notes:

- ▶ The political difficulties of working with refugees in the region and how this affects coverage, for example the Tanzanian government being unwilling to accept formal reduction to the general rations for Burundian and Congolese refugees;
- ▶ The impact of security concerns and restrictions on travel which have made access to some geographical areas problematic, e.g. in parts of Burundi;
- ▶ The ways in which the PRRO has modified its approach away from generalized relief in favor of a more targeted focus in some parts of the region;
- ▶ Targeting between different income groups within refugee camps;
- ▶ Targeting of refugees outside of camps, a population that may be missed in evaluations;
- ▶ The checklist on meeting WFP gender equality related commitments includes sex-disaggregated data on food distribution and attempts to close the gender gap in food distribution and education; and
- ▶ The need for practical guidelines to select the most vulnerable households for higher generalised rations.

Continued



## Good practice in assessment of coverage 2 continued

### WFP evaluation of food aid for relief and recovery in the Great Lakes

The report notes as will the problematic cultural issue of individual targeting (p20): ‘Targeting vulnerable individuals in a household or community in a culture where sharing is the norm can actually put that individual at personal risk when there is a shortage of food and/or other resources.’

**Source** WFP (2002) *Full Report of the Evaluation of the Great Lakes PRRO 6077 and PRRO 6077/1 Food Aid for Relief and Recovery in the Great Lakes Region*. Rome: WFP.

## 3.5 Relevance/appropriateness

### Definition

**‘Relevance** is concerned with assessing whether the project is in line with local needs and priorities (as well as donor policy).’

**‘Appropriateness** is the tailoring of humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly.’

### Explanation of definition

Relevance and appropriateness are complementary criteria which can be used at different levels. Relevance can be used to evaluate the wider elements of the intervention such as the overall goal, while appropriateness can be used to evaluate inputs and activities. Although interventions may be relevant at the macro-level, this does not necessarily mean that they are appropriate in terms of the type of activity selected. For example, improvement of nutritional status may be considered a relevant intervention, but distributing large quantities of food aid may not be the best activity to achieve this; alternatives could be food or cash for work, or measures to improve the functioning of local markets (OECD-DAC 1999). Normally, if an intervention is not relevant then it will also be inappropriate.

Additionally, the appropriateness of the actual resources or support provided should be evaluated. To continue the example above, even if food aid is considered an appropriate intervention, the type of food distributed should also be considered.

### Issues to consider

**Analysis of context and an adequate needs assessment** are of particular importance for promoting relevant and appropriate responses, and evaluators should pay close attention to the extent to which the planning, design and implementation of interventions took into account local context. Interventions are more likely to be relevant and appropriate if they have firstly carried out an adequate needs assessment, and secondly understood and supported the livelihoods and capacities of the affected population (ALNAP 2002). A needs assessment would be considered adequate if it clearly identifies, in a participatory fashion, the differentiated needs of the affected population, including how external intervention is likely to support livelihood strategies. Cultural appropriateness should also be considered; for example, an evaluation after the 1998 floods in Bangladesh found that shelter would have been more appropriate if it had been constructed with private space, including latrines, for women and girls.

Evaluators need to pay attention to **questions of cultural relativism**. For example, in countries with a relatively high standard of living, should interventions be looking to return primary stakeholders to their original condition, or to provide levels of support equal to those in responses in less developed countries (eg the Kosovo experience (ALNAP 2001), the evaluation of DFID’s response in Montserrat (DFID 1999), and the IFRC good practice case below)? There is no easy answer to the question of what constitutes ‘need’, but it is a question that evaluators should bear in mind when considering the relevance of the response. In some cases the best that evaluators can do is to flag key issues, rather than resolve them.

Of the major sectors covered in humanitarian action, **housing has often proven the most problematic in terms of relevance and appropriateness**. The first issue is of relevance,

that is whether shorter-term relief interventions should be supporting housing reconstruction as opposed to shelter, given the time, resources and expertise needed for large scale housing construction, and that high levels of primary stakeholder participation may be necessary to ensure effective programmes. In relation to appropriateness, issues have often arisen around the siting of new settlements, the method used for construction (e.g. owner construction as opposed to contractor construction), and the appropriateness of different kinds of housing design and building materials (ALNAP 2002). Evaluators may need to pay particular attention to relevance and appropriateness when evaluating housing programmes.

Evaluations should also evaluate **institutional capacity**, that is whether there is the capacity in terms of staffing, local knowledge and experience in the country or region, to make a relevant and appropriate response (see the IFRC good practice case below for further details).

## KEY MESSAGES

- ▶ Relevance and appropriateness are complementary criteria that are used to evaluate both the wider goal of the intervention and its specific approach in terms of how it responded to local context and needs.
- ▶ If lessons are to be learned from evaluations, assessment of relevance and appropriateness should involve an examination of why interventions are relevant and/or appropriate in some cases, and not in others.
- ▶ Evaluators should evaluate the extent to which the perceived needs of different stakeholders, in particular women and men, and girls and boys, in the affected population, were met by the intervention.

### Good practice in assessment of relevance/appropriateness 1

#### Evaluation of IFRC's response after the 1999 earthquake in Turkey

##### Background

Part of the evaluation of IFRC's response to the 1999 earthquake in Turkey was a thorough analysis of the relevance and appropriateness of the response in terms of: the capacity of IFRC to respond in an urban and European context, given its orientation towards working in rural developing countries; and the appropriateness of the resources provided.

Continued

## Good practice in assessment of relevance/appropriateness 1 continued

### Evaluation of IFRC's response after the 1999 earthquake in Turkey

#### Evaluation of relevance/appropriateness

In the first area, the report examines key contextual questions:

- ▶ Did IFRC have adequate competence to deal with earthquakes, in particular regarding scenario planning, seismological analysis, hazard mapping, research, contacts with specialist bodies, and framework operating agreements with National Societies in earthquake-prone countries?
- ▶ Did IFRC have the capacity to make an appropriate response in an urban setting in a European country, where there are high population densities, a heavy reliance on complex but disrupted support infrastructure, high affected population expectations, the likelihood of international media attention, and high donor interest.

The evaluation notes that, although IFRC has pointed out the humanitarian implications of a global shift towards increased urbanization at a policy level and through publications such as the *World Disasters Report*, this did not translate into a relevant response in the Turkey case. It may be useful in EHA to assess relevance against policy in this fashion.

The results of this lack of capacity are outlined in the evaluation in the assessment of the resources provided. As the evaluation notes (p83): 'the International Federation's relief items are predicated upon its many years of operationality in less developed countries, and are essentially geared to the needs of displaced populations. Many of its standard specifications dictate a basic but functional quality, which were seen as unacceptably poor by the beneficiaries in Turkey. There was also an under-estimation of the extent to which people would rescue belongings from the houses, or gain access to basic, but often superior, items through other means.'

The evaluation also deals with what might be considered a more traditional approach to assessing responses, that is examining the appropriateness of the relief items provided – tents, hygiene parcels, kerosene etc – in terms of: their actual usefulness to the affected population; whether the timing of provision was adequate; and whether they were appropriate in the context, including primary stakeholder views on this issue.

The evaluation raises the question as to whether agencies such as IFRC should be providing a global minimum standard, as defined for example by SPHERE, or whether they should provide items and resources that are appropriate to a country close to a European standard of living.

**Source** IFRC (2000) *The Turkey Earthquakes Evaluation: An independent analysis of the Red Cross/Red Crescent response to the earthquake of August and November 1999 in north-west Turkey, with recommendations*. (Geneva: IFRC)

## Good practice in assessment of relevance/appropriateness 2

### WFP Evaluation of Food Aid for Relief and Recovery in Somalia

#### Background

The evaluation assesses three year support to some 1.3 million people, with 63,000 MT of food commodities distributed at a cost of some US\$55 million. Of this support, 51 per cent was projected to go towards rehabilitation and recovery, 30 per cent to emergency relief, and 19 per cent to social institutions. The primary aim of the Protracted Relief and Recovery Operation (PRRO) was to (p4): 'contribute to a broader framework for integrated rehabilitation programmes in Somalia, while maintaining flexibility to both grasp development opportunities and respond to emergency situations.' The evaluation therefore needed to examine the relevance of this mix of allocations as well as the appropriateness of each type of intervention. The evaluation was carried out by two expatriates who visited Somalia for a three week period in mid-July 2001.

#### Evaluation of relevance/appropriateness

The overall relevance of the intervention is considered in the context of the political economy of aid in Somalia. The evaluation carefully weighs the rationale for providing food aid in Somalia. On the one hand, the country is usually in food deficit, populations in many locations are isolated from customary markets, and doubly disadvantaged due to loss of primary occupations and assets. On the other, that it may make more sense to give the affected population funds to buy local food where this is available, whether as cash-for-work or food-for-work. The analysis of the relevance of the different modes of intervention are also linked to questions of connectedness. For example, the evaluation of support to social institutions notes that (p15): 'it was appropriate for WFP to subsidise with part of its food aid the quest of local communities, nascent government departments and agencies to rebuild and run institutions, including hospitals and orphanages ... However ... these commitments of their nature tend to be longer-run than most R&R projects ... it was not clear what the exit strategy was in a number of instances.'

Lastly, the focus on relevance is complemented by an extensive discussion of the appropriateness of rations. This includes detailed analysis of the ration make-up, within the context of the Somali food economy. This evaluation's examination of both wider and specific issues means that its analysis of relevance/appropriateness is comprehensive.

**Source** WFP (2002) *Full Report of the Evaluation of PRRO Somalia 6073.00, Food Aid for Relief and Recovery in Somalia*. Rome: WFP.

## 3.6 Connectedness

### Definition

‘Connectedness refers to the need to assure that activities of a short-term emergency nature are carried out in a context that takes longer term and interconnected problems into account.’

### Explanation of definition

Connectedness has been adapted from the concept of sustainability - the idea that interventions should support longer-term goals, and eventually be managed without donor input. Although it is generally accepted that there is always a link between humanitarian action, reconstruction and development, and that humanitarian action should establish a framework for recovery, there is currently no consensus concerning the *extent* to which humanitarian action should support longer-term needs. This may make evaluation of connectedness problematic, in particular in complex emergencies where there is limited development activity, or in natural disasters where primary stakeholders are semi-permanently caught in the relief phase. Evaluators should concentrate in particular on whether the **key linkages between the relief and recovery phases** have been established, for example the existence of a sound exit strategy with timelines, allocation of responsibility and details on handover to development agencies and/or government departments, and adequate availability of funding post-response.

### Issues to consider

**Evaluators should be sensitive to the way in which funds have been spent and evaluate the intervention accordingly.** Substantial amounts of humanitarian aid are regularly spent on reconstruction or rehabilitation, rather than relief (ALNAP 2002). Rehabilitation/ reconstruction should be evaluated against appropriate indicators, which may not be the same indicators used to evaluate relief (e.g. mouths fed or lives saved). However, attempts to do this may be hindered by inadequate financial records, and objective statements in planning documents, which are not always sufficiently clear.

**Connectedness and partnerships.** Humanitarian action tends to promote connectedness more effectively where partnerships, particularly between international and national NGOs, already exist in the development sphere. Evaluators should analyse the nature of partnerships supporting connectedness, e.g. how they came into being and were supported, so that others can learn from this experience.

**Evaluations should examine the extent to which local capacity is supported and developed.** For example, a number of evaluations have pointed out that establishing capacity of water-user associations to manage water facilities may be too great a challenge for the relief phase (see the ECHO good practice example and ALNAP 2002). Evaluations should also examine the degree to which livelihoods of the affected population are supported or disrupted by the intervention, as this will have a significant impact on longer term results. They should also analyse the degree to which the capacity of government at various levels is built by the intervention.

**In the LFA,** details on connectedness will usually be found at the purpose or goal levels; the assumptions column may also include information about risks related to connectedness.

## KEY MESSAGES

- ▶ Large-scale relief programmes can have a significant long-term impact, for example on local power structures; on government capacity; on gender equality; or the environment. It is important for evaluators to examine each of these, and any other relevant areas.
- ▶ Evaluators should determine whether funds have been used for relief or reconstruction, and evaluate accordingly.
- ▶ When evaluating connectedness, evaluators should pay particular attention to institutional factors, specifically the existence of strong partnerships, and the extent to which national/local capacity is supported and developed.

### Good practice in assessment of connectedness 1

#### Evaluation of ECHO's Global Plans

##### Background

The ECHO Manual for the evaluation of humanitarian aid requires an assessment of connectedness – which it terms ‘viability’ – in its commissioned evaluations. LRRD is a usually a central theme in evaluations of ECHO funded interventions. The Manual states (1999: 12): ‘Humanitarian aid must build on local capacities, reinforcing coping mechanisms and institutions. It is a basic philosophy that every possible step must be taken to prevent the beneficiaries of humanitarian aid from becoming dependent on it, and that self-sufficiency should be the goal.’ ECHO evaluations provide two examples of good practice in the evaluation of connectedness.

##### Evaluation of connectedness

One part of the evaluation of ECHO's Global Plan in Angola is a report on three water and sanitation interventions implemented by NGOs. The evaluation succinctly notes that although in each case the water-related intervention met its short term objectives, it was unlikely to contribute to longer term development. As it notes concerning one intervention (2001: 13): ‘This project can be best classified as a successful emergency recovery programme.... However, a lasting impact on the water supply is not guaranteed, since the project did not develop a long-term pump maintenance strategy...it is known from other countries that the maintenance issue of hand pumps is the essential factor in any rural water supply strategy...It is unrealistic to expect that newly introduced hand pumps within the framework of an emergency project will survive long after the end of the project, even when training is given and spare parts have been supplied.’ Connectedness may be a particular problem in the water sector, where it has often proven difficult to establish longer-term solutions to water management (ALNAP 2002).

Continued

## Good practice in assessment of connectedness 1 contd

### Evaluation of ECHO's Global Plans

The evaluation of ECHO's support to the health sector in Central America makes similar points, in this case in relation to unsustainable institutions. This evaluation raises important questions concerning how far it is possible for an emergency intervention to support longer term development, questions that are often rightly in the back of evaluators' minds, and the overlapping nature of relief and rehabilitation work. It notes that (2001a: 10): 'Around 90% of all people interviewed confided that almost all health programmes financed by ECHO would collapse shortly after the INGO would stop its support.' The report highlights the problem of the lack of connection between ECHO and EC development departments, which has been noted in a number of ECHO evaluations (ALNAP 2001).

**Sources** ECHO (2001) *Evaluation of ECHO's Global Plan 2000- Angola. Evaluation Report 305*. (Brussels: ECHO)

ECHO (2001a) *Central America. Hurricane Mitch Global Plans 1998 and 1999. Health Sector-2. Evaluation Report 300*. (Brussels: ECHO)

## Good practice in assessment of connectedness 2

### Evaluation of Disasters Emergency Committee Earthquake Appeal Funds

#### Background

The Disasters Emergency Committee (DEC) evaluation of British NGO interventions after the earthquake in Gujarat in January 2001 is an unusually detailed report resulting from collaboration between eight specialists from three organisations based in India and the UK, and based on three visits by the UK team members between March and October 2001. The evaluation used the Red Cross Code as the set of values against which the response was measured, and also included a public opinion survey covering 50 villages and some 2,300 people.

#### Evaluation of connectedness

This evaluation illustrates the ways in which connectedness needs to be viewed from macro and micro perspectives, from the level of national level institutions to the level of local livelihoods.

Continued



## Good practice in assessment of connectedness 2 contd

### Evaluation of Disasters Emergency Committee Earthquake Appeal Funds

The evaluation examines the key area of timing, and that few agencies made a strategic review at the crucial point when turning from relief to rehabilitation. It analyses as well the quality of partnership between international and local NGOs, the key role that good partnerships and local capacity played in successful interventions, and whether agencies decided to work with expatriate or local staff.

At a general level, the evaluation makes the point that NGOs directly intervening after the earthquake may have been better utilizing their resources towards pressurizing the various levels of government to direct resources to the most vulnerable groups, as the political context meant that NGO support likely replicated rather than replaced what the government would have in any case done. The potential to bring about social change through advocacy was therefore lost. In this case therefore attention to connectedness involved not so much building local capacity as emphasising the links between external and internal actors.

At the level of specific interventions, the evaluation includes a detailed analysis of support to rehabilitation and livelihoods. In terms of the impact of connectedness on the affected population, the area of livelihood support is a key one that is often missed in EHA. It considers the tradeoffs in water harvesting between speed and quality, that is between whether to build as many water control structures as possible or ensure the greatest community mobilization. In the area of housing the evaluation examines the connection between shelter and livelihoods and whether the housing reconstruction strategy chosen by NGOs was likely to lead to sustainable solutions for the affected population.

**Source** DEC (2002) *Independent Evaluation of Expenditure of DEC India Earthquake Appeal Funds*. London: Disasters Emergency Committee

## 3.7 Coherence

### Definition

‘The need to assess security, developmental, trade and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations’

### Explanation of definition

The evaluation of coherence complements that of coordination. While assessment of coordination tends to focus on operational issues, assessment of coherence should focus more on the extent to which policies of different actors were complementary or contradictory.

Coherence becomes an important evaluation issue when politics fosters the occurrence or continuation of a humanitarian emergency; and when military and civilian actors are involved in the same emergency – for example when humanitarian actors are denied access to certain regions by the military for security reasons. Coherence can also be analysed solely within the humanitarian sphere – to assess whether all actors – including governments, the UN, donors, civil society and the private sector – are working towards the same goals.

As the definition makes clear, evaluation of coherence needs to take into account considerations of humanitarian space, including protection. For example, there have been instances of one UN agency promoting the return of refugees to their host country while another is opposed to such an approach (OECD-DAC 1999). Evaluations need to consider whether actors have been coherent in their approach to protection and whether policy has met the protection needs of primary stakeholders.

Coherence has proven the most difficult of the DAC criteria to operationalise (ALNAP 2003). It is often confused with ‘coordination’; in addition, most EHA does not consider the policy of the agency being evaluated, let alone the policies of multiple agencies. It may be less relevant to consider coherence in single agency, single project evaluations, but if an evaluation is not going to consider policies as benchmarks against which to measure results, the reasons for this should be made clear.

### Issues to consider

Evaluation managers need to ensure **that evaluation teams have the capacity and resources to evaluate coherence**, in particular specialists in policy analysis or the evaluation of civil-military relations. They also need to be able to live with the consequences of evaluating coherence, because it is the most ‘political’ of the DAC criteria, given its focus on wider policy issues. The most notable recent example is the evaluation of the Great Lakes emergency in 1994, where military contingents were withdrawn from Rwanda during the crisis, despite evidence to suggest that a rapid deployment of troops could have prevented many of the killings and subsequent refugee influx into Zaire, leading to a massive humanitarian response (OECD-DAC 1999). The political ramifications of evaluation findings were felt far and wide, including in the upper reaches of the UN and several governments.

Since the early 1990s, **military forces have been increasingly involved in humanitarian assistance**, in some cases supporting humanitarian agencies and in a few cases providing aid

directly, and this trend seems likely to continue. Appropriate levels of collaboration and roles for humanitarian actors and military forces are still much debated. Barry and Jefferys (2002) argue that the military has a core mandate to foster security and protect civilians, while humanitarian agencies have a mandate to implement humanitarian assistance impartially, and that these activities should be kept separate.

Because of the increased involvement of military forces, evaluators need to pay close attention to the mandates, missions and principles of different actors, and evaluate how far these mandates contradict or complement each other.

**Some NGOs are increasingly becoming involved in advocacy work to complement their operational activities** (e.g. ActionAid 2001). In emergencies where there are a number of actors, including the government and the military, evaluations need to assess whether NGOs should have undertaken an advocacy role, and for those that did this, how appropriate and effective it was. The evaluation of the DEC agencies response in Gujarat for example (DEC 2002) criticized the majority of the DEC agencies because they did not sufficiently attempt to influence the response of the Governments of India and Gujarat.

---

## KEY MESSAGES

---

- ▶ Evaluation of coherence will always be politically charged, more so than with use of the other DAC criteria. It may also be the most difficult of the DAC criteria to evaluate, in particular in single agency, single project evaluations.
- ▶ Evaluating coherence is of particular importance when there are a number of actors involved in the response, as they may have conflicting mandates and interests.
- ▶ Important questions to ask in the case of coherence are: why was coherence lacking or present; and, what were the particular political factors which led to coherence or its lack?

<sup>8</sup>There are a very limited number of examples of good practice of the evaluation of coherence and only one example is provided here.

## Good practice in assessment of coherence<sup>8</sup>

### Kosovo: evaluation of UNHCR's emergency preparedness and response

#### Background

The international response to the 1999 Kosovo conflict ranks as one of the largest, in terms of the scale of resources involved, and this in turn generated a multitude of evaluation reports. The focus and quality of these evaluation reports is the main focus of the ALNAP *Annual Review 2001*.

As the *Annual Review* points out, coherence is one of the central evaluative issues of the Kosovo conflict (ALNAP 2001: 72): 'What distinguishes this particular humanitarian action from many others is the extent to which it is dominated by the dilemmas and paradoxes thrown up by NATO's involvement – particularly since those governments sending in bombers were also funding humanitarian efforts. Programmes and evaluations alike recognize that NATO's involvement in the overall sequence of events was huge and decisive.' Given NATO's significant role in both military and humanitarian action (e.g. refugee camp construction), it might have been expected that evaluation reports would have focused on the issue of coherence, particularly as the bombing was being carried out without UN sanction. In fact, few of the evaluations managed a consistent or analytical focus on coherence. One of the reasons for this may be that most of the Kosovo evaluations were single agency, and there is no system-wide evaluation which might have produced greater attention to coherence.

#### Evaluation of coherence

The evaluation of UNHCR's response was one of the exceptions in terms of its systematic attention to coherence. One chapter is dedicated to relations with the military; the chapter outlines the ways in which UNHCR's cooperation with NATO in Kosovo was a departure for the agency, given that NATO was a party to a conflict unauthorized by the UN Security Council. It goes on to analyse the policy aspects of NATO-UNHCR relations through the lens of the most visible and concrete forms of cooperation. The evaluation also covers policy coherence as far as protection of refugees was concerned.

The areas of cooperation agreed between the High Commissioner for Refugees and the Secretary-General of NATO were logistics (airlift operations, offloading and storage of aid), construction of refugee camps, UNHCR facilitating agreement from its member states to take some refugees from FYR Macedonia, and help in transporting them to third countries. The policy implications of each of these areas of cooperation is outlined from the perspective of both partners.

## Good practice in assessment of coherence continued

### Kosovo: evaluation of UNHCR's emergency preparedness and response

The evaluation also covers one of the current key issues in the assessment of coherence – how far the line dividing the military and humanitarian spheres is maintained. In the Kosovo case this line was blurred, with some negative consequences for refugees, for example some camps being located too close to the theatre of war. The evaluation also discussed the implications of military-humanitarian cooperation for the longer term work of UNHCR; one of the implications appears to be the need for greater clarity at the policy level by humanitarian agencies as to what the relationship should constitute.

Source: UNHCR (2000) *The Kosovo Refugee Crisis: An independent evaluation of UNHCR's emergency preparedness and response*. Geneva: UNHCR.

<sup>9</sup> All of the criteria will be useful in most evaluations to some extent. This column selects evaluation types where the criteria will be particularly useful.

Table 3.1 Summary of main information on DAC criteria

Criteria	Definition	Main use <sup>9</sup>
<b>Effectiveness</b>	A measure of the merit or worth of an activity, i.e. the extent to which a development intervention has attained or is expected to attain, its relevant objectives efficiently and in a sustainable way	Single sector or single agency evaluations
<b>Impact</b>	The totality of positive and negative, primary and secondary effects produced by a development intervention, directly or indirectly, intended or unintended	Multi-sector, multi-donor evaluations; system-wide evaluations
<b>Efficiency</b>	Efficiency measures outputs — qualitative and quantitative — in relation to inputs. This generally requires comparing alternative approaches to achieving the same output, to see whether the most efficient process has been used	All evaluation types where adequate financial information is available
<b>Coverage</b>	The need to reach major population groups facing life-threatening suffering wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agendas	All evaluation types except those with a mainly institutional focus
<b>Relevance/ Appropriateness</b>	The extent to which the objectives of a development intervention are consistent with country needs, global priorities and partners' and donors' policies	Policy based evaluations, all other evaluation types except those with a mainly institutional focus
<b>Connected-ness</b>	The extent to which the activities take account of longer term needs and interconnectedness of humanitarian problems	Evaluations assessing institutional structures and partnerships
<b>Coordination</b>	The systematic use of policy instruments to deliver humanitarian assistance in a cohesive and effective manner	Institutional evaluations are likely to review strategic coordination; operational coordination can be evaluated in most other types of evaluations
<b>Coherence</b>	The need to assess security, developmental, trade and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations	Policy-based evaluations

## 4 References

- ActionAid (2001) *ActionAid Emergencies Impact Assessment Study*. London: ActionAid
- ALNAP (2002) *Humanitarian Action: Improved performance through improved learning*. ALNAP Annual Review 2002. London: ALNAP/ODI.
- ALNAP (2001) *Humanitarian Action: Learning from Evaluation*. ALNAP Annual Review 2001. London: ALNAP/ODI.
- Bamberger, M. (2000) *Integrating quantitative and qualitative research in development projects*. Washington D.C.: World Bank.
- Barry, J. with A. Jefferys (2002) *A bridge too far: aid agencies and the military in humanitarian response*. London: ODI Humanitarian Practice Network Paper 37.
- Broughton, B. (2001) *How Logframe approaches could facilitate the planning and management of emergency operations*. Canberra: mimeo.
- DEC (2001) *Independent Evaluation of DEC Mozambique Floods Appeal funds: March 2000-December 2000*. London: Disasters Emergency Committee.
- DEC (2000) *Independent Evaluation of Expenditure of DEC Kosovo Appeal Fund: Phases I and II, April 1999-January 2000, Volumes I, II and III*. London: Disasters Emergency Committee.
- ECHO (1999) *Manual for the Evaluation of Humanitarian Aid*. Brussels: ECHO.
- Fetterman, D. (2002) 'Empowerment Evaluation: Building Communities of Practice and a Culture of Learning.' *American Journal of Community Psychology* (30), 1, 89-102.
- Frohardt, M., Paul, D. and L. Minear (1999) *Protecting Human Rights: The Challenge to Humanitarian Organizations*. Providence: Humanitarianism and War Project, Occasional Paper 35.
- Groupe URD (2000) *NGO Responses to Hurricane Mitch: Evaluations for accountability and learning*. London: ODI.
- Hallam, A. (1998) *Evaluating Humanitarian Assistance Programmes in Complex Emergencies*. London: ODI Good Practice Review 7.
- IFRC (2002) *Operational Framework for Evaluation*. Geneva: IFRC.
- IRC (2001) *The IRC Causal Pathway Framework. A Guide to Program Design, Monitoring and Evaluation*. New York, International Rescue Committee.
- Kanbur, R. (ed) (2001) *Qualitative and Quantitative Poverty Appraisal: Complementarities, Tensions and the Way Forward*. Contributions to a Workshop held at Cornell University. <http://people.cornell.edu/pages/sk145/papers/QQZ.pdf>
- Lackey, Jill et al (1997) "By Whose Standards? Reflections on Empowerment Evaluation and Grassroots Groups." *Evaluation Practice* 18 (2), 137-46.
- Mosse, D. J. Farrington and A. Rew (1998) *Development as process. Concepts and methods for working with complexity*. London: Routledge.
- ODI (2003) *According to need? Needs assessment and decision-making in the humanitarian sector. Summary report*. London: Humanitarian Policy Group Report 15.
- OECD-DAC (2000) *DAC Criteria for Evaluating Development Assistance*. Paris: OECD.
- OECD-DAC (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*. Paris: OECD.
- OECD-DAC (1991) *Principles for Evaluation of Development Assistance*. Paris: OECD
- Patton, M. (1997) *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks: Sage.

- Paul, D. (1999) *Protection in Practice: Field-Level Strategies for Protecting Civilians from Deliberate Harm*. London: ODI Relief and Rehabilitation Network Paper 30.
- Preskill, H. and V. Caracelli (1997) “Current and Developing Conceptions of Use: Evaluation Use TIG Survey Results” *Evaluation Practice Fall 1997* 18.3.
- Reindrop, N. and P. Wiles (2001) *Humanitarian Coordination: Lessons from Recent Field Experience. A study commissioned by the Office for the Coordination of Humanitarian Affairs (OCHA)*. Geneva: OCHA.
- Shulha, Lyn and B. Cousins (1997) “Evaluation Use: Theory, Research, and Practice since 1986.” *Evaluation Practice* 18 (3), 195-208.
- SPHERE (2000) *Humanitarian Charter and Minimum Standards in Disaster Response*. Geneva: The Sphere Project.
- Stufflebeam, D. (1999) *Foundational Models for 21st Century Program Evaluation*. Western Michigan University, Evaluation Centre Occasional Paper Series.
- UNHCR (2002) *Real-time humanitarian evaluations: some frequently asked questions*. Geneva: UNHCR.
- UNHCR (2002a) *Meeting the rights and protection needs of refugee children. An independent evaluation of the impact of UNHCR's activities*. Geneva: UNHCR.
- UNHCR (2001) *Step by Step Guide to Managing Evaluations*. Geneva: UNHCR, draft.
- UNHCR (2000) *The Kosovo Refugee Crisis: An independent evaluation of UNHCR's emergency preparedness and response*. Geneva: UNHCR.
- UNICEF (2001) *Managing monitoring and evaluation activities. Monitoring and evaluation in crisis and unstable contexts*. New York: UNICEF.
- Valadez, J. and M. Bamberger (1994) *Monitoring and Evaluating Social Programs in Developing Countries*. Washington D.C.: World Bank Economic Development Institute Development Series.
- Van Brabant, K. (1999) *Opening the Black Box. An outline of a framework to understand, promote and evaluate humanitarian coordination*. London: Overseas Development Institute.
- van de Putte, B. (2001) “Follow up to evaluations of humanitarian programmes”. Paper submitted to the ALNAP Biannual Meeting, 26-27 April.



# Annex 1 American Evaluation Association Guiding Standards for Evaluation

To improve performance, EHA needs to be contextualized within wider good practice in the evaluation field in general. The American Evaluation Association (JCSEE 1994) has developed what are currently the most comprehensive standards for sound evaluation – specific to evaluations of educational initiatives, but which can be adapted to guide EHA. As these standards are in widespread use in the evaluation community, it makes sense to employ them to guide understanding of ways of working with the DAC criteria, and discussion of the standards is integrated into this Handbook as relevant.<sup>10</sup>

The central points of relevance to EHA of the four basic evaluation standards – utility, propriety, feasibility, and accuracy – are summarized below after relevant components of the standards are noted. The first standard in particular – utility – is a central topic which cross-cuts this Handbook, as clearly is evaluations are not used there is little or no point in conducting them. One of the questions this Handbook asks is: how should the DAC criteria be reformulated to concentrate on evaluation use? Not all of the standards are quoted below, and can be found at: <http://www.eval.org/Publications/publications.html#EDUCATIONAL%20EVALUATION%20STANDARDS>

## Utility

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

- **Stakeholder Identification.** Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.
- **Values Identification.** The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.
- **Report Timelines and Dissemination.** Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.
- **Evaluation Impact.** Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

**Relevance to EHA** Like other evaluation practice, EHA needs to pay close attention to intended uses of the evaluation. Terms of reference should set out the ways in which the commissioning agency is planning to use evaluation results. Evaluation workplans should build in mechanisms for regular feedback sessions with key stakeholders. Evaluators should be proactive in

<sup>10</sup> DAC (1991) also developed a set of principles for evaluation of development, reviewed in 1998 (DAC 1998) which overlap to a certain extent with the standards discussed in this Section. As well as the Evaluation Standards, the American Evaluation Association has developed a set of Evaluation Principles which can be found at <http://www.eval.org/EvaluationDocuments/aeaprin6.html>

promoting evaluation use, investigating how and why results from previous evaluations have been used in commissioning agencies.

## Propriety

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

- **Rights of Human Subjects.** Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.
- **Human Interactions.** Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.
- **Complete and Fair Assessment.** The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

**Relevance to EHA** In both complex emergencies and natural disasters, evaluators should ensure confidentiality for key stakeholders who may be at risk of recrimination. All respondents should be treated with respect, whatever their political or religious affiliation or social status. The focus on outlining strengths and weaknesses, now quite common in EHA, should include an analysis of related ameliorative action.

## Feasibility

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

- **Political Viability** The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

**Relevance to EHA** Evaluators need to make themselves aware of interest groups in the commissioning agency and what their interest in the evaluation is likely to be. Any potentially controversial conclusions and recommendations need to be justified as clearly as possible.

## Accuracy

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth of merit of the program being evaluated.

- **Program Documentation.** The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

- **Context Analysis.** The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.
- **Described Purposes and Procedures.** The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.
- **Defensible Information Sources.** The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
- **Valid Information.** The information gathering procedures should be chosen or developed and then implemented to assure that the interpretation arrived at is valid for the intended use.
- **Reliable Information.** The information gathering procedures should be chosen or developed and then implemented to assure that information obtained is sufficiently reliable for the intended use.
- **Justified Conclusions.** The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.
- **Impartial Reporting.** Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

**Relevance to EHA** Currently the majority of EHA does not include adequate information for most of the accuracy points. EHA needs to pay much more attention to factoring in the importance of contextual issues, describing and justifying methods used, and sources and analysis of information gathered (ALNAP 2002).

Among the implications of applying these general evaluation standards to guide EHA is a move away from the idea that the evaluator is an independent “objective” observer, as is currently explicit or implicit in many evaluations of humanitarian action. Use of the standards implies that the evaluator is a key player within an often highly charged political environment, and within this environment the evaluator needs to promote, through proactive planning, the use of evaluation results.

## Annex 2 Main types of evaluation of humanitarian action

<b>System-wide</b>	Conducted jointly with the participation of all or most donors involved, although may be led by one or two donors, and covering the total intervention. May also include the host government(s)
<b>Joint</b>	Conducted jointly with other donors, but not including all involved. Focuses on interventions of agencies carrying out evaluation
<b>Synthesis</b>	Synthesis from a number of evaluations of individual programmes
<b>Single Agency</b>	All interventions of one agency
<b>Policy</b>	Whether policy has been followed. Alternatively, may involve evaluation of the framework of understanding, beliefs and assumptions that direct individual programmes
<b>Institutional/process</b>	The internal dynamics of the implementing organization, its policy instruments, service delivery mechanisms, management practices and linkages among these
<b>Thematic</b>	One theme, such as gender or environment, across a number of programmes or an organization as a whole
<b>Sectoral</b>	Aid to a particular sector, such as health or water, in a number of projects or country programmes

Adapted from ALNAP Training Module 1, [www.alnap.org](http://www.alnap.org)

# Annex 3 Good practice in methods for the Evaluation of Humanitarian Action

EHA differs from general evaluation as well as evaluation of development projects, but it retains in common with these types of evaluation the need to be as rigorous and credible as possible. Below are some pointers to good practice in development of methodology for EHA.<sup>11</sup>

**Ensure that the method to be used is adequately described** in the terms of reference, and that the method actually used is adequately described in the evaluation report. It is common for evaluations to state in a few lines or a short paragraph the method used, which undermines the credibility of reports. If readers don't like the findings of an evaluation, the first thing to be questioned is the method! Evaluations should in particular note: reasons for choice of geographical locations and projects visited; who was interviewed, why, and in what setting; and any constraints faced.

**Use a multi-method approach, and cross-check wherever possible** In development evaluations, this tends to mean combining structured and semi-structured questionnaires, the results of which are often analysed statistically, with more participatory approaches such as Participatory Rural Appraisal. In EHA there is usually little scope for selecting samples and carrying out statistical analysis. Instead, data which is available in government or agency reports, for example on numbers and types of primary stakeholders covered, can be cross-checked in interviews with agency staff and primary stakeholders.

**Assess the intervention against appropriate international standards and law** Standards such as UN Conventions, the Red Cross/Red Crescent Code of Conduct and Sphere should be routinely used as part of the evaluation methodology.

**Talk to primary stakeholders** There are two main reasons for doing this. Firstly there is an ethical requirement that evaluators do their best to talk to as many primary stakeholders as possible to ensure that the interests of this group are adequately represented in the evaluation. This also ensures accountability of evaluators to primary stakeholders. Secondly, EHA should be as comprehensive as possible and if the views of the main group that is intended to benefit from humanitarian action are not represented, evaluation credibility is undermined. Evaluations need to be as representative as possible, so efforts should be made to talk to both sexes, primary stakeholders from different socio-economic groups and ethnicity, and adults and children. Evaluators need to be sensitive however to the past experience of primary stakeholders; while some may be keen to talk to evaluators, others may not wish to relive their past experiences simply for the sake of an evaluation. Like any other source of data, the perspectives of primary stakeholders should be cross-checked.

**Disaggregate** Evaluations are stronger where information is broken down by sex, socio-economic group and ethnicity. Disaggregated data may not always be available in government and agency reports, in which case the onus is on the evaluation team to supplement existing data.

<sup>11</sup> There are a number of good sources for use of methods in EHA. The ALNAP Training Module (part 2) gives a general background. The ALNAP Proforma is a checklist of good practice for both development of method and EHA more generally and can be found in the *ALNAP Annual Review*. Both of these can be found at [www.alnap.org](http://www.alnap.org). Hallam (1998) is a further useful source.

**Ensure there is a focus on social process and causality** EHA tends to concentrate on what happened rather than why it happened, which does not support lesson learning to the extent possible. Evaluators need to build into their evaluation questionnaires questions such as: ‘What were the reasons the intervention succeeded or failed?’ in order to support understanding of social process and lesson learning.

**Make clear any evaluator bias** All evaluators bring their personal biases (for example, a desire to empower primary stakeholders is one kind of bias, or a belief in ‘objective’ evaluation is another). If these biases are made clear then the evaluation will be more credible. Using multiple sources of information and cross-checking usually helps cut down evaluator bias.

Last but not least – **integrate the DAC criteria**, or provide reasons for not using the criteria!