

Evaluating humanitarian action using the OECD-DAC criteria

**An ALNAP guide for
humanitarian agencies**

ISBN 0-85003-795-6

Overseas Development Institute ©

London, March 2006

Photocopies of all or part of this publication may be made providing that the source is acknowledged. ALNAP would appreciate receiving details of any use made of this material in training, research or programme design, implementation or evaluation. Requests for commercial reproduction of ALNAP material should be directed to ALNAP.

For further copies of this publication, please contact:

ALNAP
c/o Overseas Development Institute
111 Westminster Bridge Road
London SE1 7JD
UK

Tel: +44 (0)207 9220300

Fax: +44 (0)207 9220399

Email: alnap@odi.org.uk

Website: www.alnap.org

Price per copy: £7.50

Foreword

The *ALNAP Review of Humanitarian Action* has been monitoring the quality of evaluations of humanitarian action for the past five years. The positive news is that evaluations have become more deeply ingrained in the humanitarian sector. The not-so-good news is that the quality of the evaluations themselves still leaves much to be desired. The development of better evaluations is not just an end in itself: high-quality evaluations should have a positive effect on improving learning, accountability and performance in the sector.

When the Joint Evaluation of Emergency Assistance to Rwanda was undertaken in 1995, there were no manuals, guidelines or good practice notes to follow on evaluating humanitarian action. Since then, some helpful materials have been published – notably *Evaluating Humanitarian Assistance Programmes in Complex Emergencies* (Hallam, 1998) and *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* (OECD-DAC, 1999).

The first of these texts was aimed at evaluators, while the second was intended for commissioning agencies. This new ALNAP guide builds on the achievement of both publications by providing a framework designed to assist with the interpretation of key DAC criteria within a specifically humanitarian context. This will be of value to both evaluators and evaluation managers alike.

Evaluation teams in the field piloted draft versions of this guide, and the practical lessons learned from those experiences have been incorporated into this final version. I believe that this process has helped to create a guide that is user-friendly and will be an excellent tool for designers of evaluations and for the teams on the ground.

Alongside other activities to improve evaluation, such as the ALNAP meta-evaluation and ALNAP evaluation training modules, this guide signals how far the evaluation community has come in pursuing quality in evaluation as a means of improving humanitarian performance. With other recent initiatives on evaluation quality and standards by both DAC and the UN, there are good grounds for expecting further improvements in the future.

Niels Dabelstein

Head of Evaluation Department

Danida

Contents

Foreword	3
Acknowledgements	7
Acronyms and abbreviations	8
1 Purpose and use of this guide	9
1.1 Background	10
1.2 Intended users of this guide	11
1.3 Organisation of the guide	12
2 Thinking about EHA	13
2.1 What is the evaluation of humanitarian action?	14
2.2 What is different about EHA?	15
2.3 The use of evaluation results	16
3 Working with the DAC criteria	17
3.1 Relevance/appropriateness	22
3.2 Connectedness	27
3.3 Coherence	33
3.4 Coverage	38
3.5 Efficiency	44
3.6 Effectiveness	49
3.7 Impact	56
4 Methods for the evaluation of humanitarian action: pointers for good practice	63
References	67
Annexe 1 Joint Committee on Standards for Educational Evaluation	71
Annexe 2 Checklist of themes to be covered in EHA	77

Tables and boxes

Box 1 What are the DAC criteria?	10
Box 2 Evaluation of humanitarian action: the ALNAP definition	14
Box 3 Red Cross/Red Crescent Code of Conduct and coverage	39
Table 1 Summary definitions of the DAC criteria	20

Good practice examples

Assessment of relevance/appropriateness

1 Evaluation of IFRC's response after the 1999 earthquake in Turkey	24
2 WFP evaluation of food aid for relief and recovery in Somalia ..	26

Assessment of connectedness

1 Evaluation of ECHO's global plans	29
2 Evaluation of DEC India earthquake appeal funds	31

Assessment of coherence

1 Evaluation of UNHCR's emergency preparedness and response, Kosovo	36
---	----

Assessment of coverage

1 Evaluation of Danish humanitarian assistance to Sudan, 1992–1998	41
2 WFP evaluation of food aid for relief and recovery in the Great Lakes	42

Assessment of efficiency

1 Evaluation of DEC Mozambique flood appeal funds	46
2 UNHCR evaluation of the Dadaab firewood project in Kenya	48

Assessment of effectiveness

1 DFID evaluation of support to WFP in Bangladesh	52
2 WFP evaluation of food assistance and support for repatriation of Iraqi and Afghan refugees in Iran	53

Assessment of impact

1 Joint evaluation of emergency assistance to Rwanda	58
2 Evaluation of ECHO health, nutrition, water and sanitation in Sierra Leone	59

Acknowledgements

This handbook was produced with the guidance and support of the ALNAP Secretariat, and in particular John Borton, John Mitchell, Sera Orzel, Kate Robertson and Gabriele Russo. 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, Samuel Taddesse, John Telford and Peter Wiles.

Acknowledgements are also due to: Peter Cavendish, Evaluation Head of Sector, ECHO, who supported the piloting of the guide in Zimbabwe and Sri Lanka during ECHO evaluations; the ECHO evaluation teams in those two countries, and in particular Richard Ellert; and Francois Grunewald of Groupe URD, for piloting the guide during an evaluation mission to Sierra Leone and Guinea.

Tony Beck

Acronyms and abbreviations

ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
DAC	Development Assistance Committee
DEC	Disasters Emergency Committee
ECHO	European Community Humanitarian Office
EHA	evaluation of humanitarian action
IDP	internally displaced person
IFRC	International Federation of the Red Cross and Red Crescent
JEEAR	Joint Evaluation of Emergency Assistance to Rwanda
JCSEE	Joint Committee on Standards for Educational Evaluation
logframe	logical framework
LRRD	linking relief, rehabilitation and development
OECD	Organisation for Economic Cooperation and Development
PRRO	protracted relief and recovery operation
UN	United Nations
WFP	World Food Programme

section one

Purpose and use of this guide

1.1 Background

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), becoming a set of seven criteria that will also be covered in this guide: relevance/appropriateness, connectedness, coherence, coverage, efficiency, effectiveness, and impact. The DAC criteria are intended to be a comprehensive and complementary set of measures.

This guide was developed after discussions within the evaluation community, and within ALNAP in particular, about how to strengthen evaluation of humanitarian practice, and how to 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) – including within evaluations themselves and as part of agency guidance. However, several criteria are not well understood; their use is often mechanistic, and excludes more creative evaluation processes.

The **objective of this guide** is to provide practical support on how to use the DAC criteria in EHA. In order to maintain focus, the guide does not include significant detail on wider issues such as evaluation management, choice of methodologies or evaluation approaches. Also, the guide is not intended to provide in-depth details on the DAC criteria but rather to introduce the criteria to the reader, who can then follow up as appropriate.

This guide draws on good-practice material on evaluation and on EHA, including other guides, handbooks and manuals (as listed under References below). The intention was to build on rather than to replicate what already exists. It also draws on a questionnaire completed by 25 ALNAP Full Members, and on the work of the author on the ALNAP Review of Humanitarian Action.

ALNAP piloted earlier versions of the guide during three humanitarian evaluations in 2004 – in Sierra Leone/Guinea, Zimbabwe and Sri Lanka, with the author joining the evaluation team in the last country. While this lengthened the process of production of the guide by a year,, it also meant that the guide was ‘road-tested’ by practitioners, and their useful feedback has been incorporated into this final version.

1.2 Intended users of this guide

This guide is intended primarily for:

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

Other intended users include developers of policy and programme design, and those involved in programme implementation. These groups have expressed a need for better understanding of the criteria against which interventions are being assessed. These interests are

addressed mainly through discussion of the links between use of the Logical Framework Analysis (logframe) and the DAC criteria, in Section 3 below.

1.3 Organisation of the guide

- **Section 2** focuses on some key themes and issues current in EHA, particularly in relation to lesson-learning and accountability, and evaluation use, to provide general context.
- For each of the seven DAC criteria, **Section 3** provides a definition, an explanation of the definition, issues to consider, key messages, and two examples of good practice intended to provide pointers and suggestions on how to use each criterion.
- **Section 4** provides brief guidelines for good practice in methods for the evaluation of humanitarian action, included at the request of several evaluators piloting the guide.

While it was decided to keep the guide tightly focused on the DAC criteria, brief details in related areas are included as annexes, in response to requests from reviewers, and from evaluators piloting the guide.

- **Annexe 1** looks at the relevance to EHA of the most commonly used set of evaluation standards (those of the Joint Committee on Standards for Educational Evaluation (JCSEE)).
- **Annexe 2** discusses eight of the main cross-cutting themes which should be included when using the DAC criteria.

section two

Thinking about EHA

2.1 What is the evaluation of humanitarian action?

This guide uses the ALNAP definition, given in Box 2. Definitions related to humanitarian action tend to stress that evaluations are objective or impartial exercises intended to promote accountability and lesson-learning. This guide is intended to support most types of evaluation.

Box 2 Evaluation of humanitarian action: the ALNAP definition¹

Evaluation of humanitarian action (EHA) is a systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability. EHA:

- is commissioned by or in cooperation with the organisation(s) whose performance is being evaluated;
- 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;
- assesses policy and/or practice against recognised criteria (eg, the DAC criteria);
- articulates findings, draws conclusions and makes recommendations.

¹ In determining this definition, there was considerable discussion among ALNAP Full Members concerning the potential for evaluations to be 'impartial'. The question of how far evaluators can or should separate their own perspective from the evaluation

2.2 What is different about EHA?

The evaluation of humanitarian action is just one type of evaluation but does have some distinct characteristics. These need to be taken into account when planning to use the DAC criteria.

Humanitarian evaluations are often undertaken during periods of severe disruption, which in the case of complex emergencies can be prolonged. Getting 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.
- Humanitarian action takes place in disordered conditions, leading to rapid change in circumstances that make context difficult to be sure of later, and also meaning that many assumptions about 'normal' social and physical conditions may no longer be justified.

Because of these constraints, when completing an evaluation report, evaluators should make clear the constraints they faced and how these constraints affected the evaluation process and findings.

process is one of the most important current debates on evaluation practice. This guide cannot cover this topic in detail, but interested readers are directed to House (2004), Lackey et al (1997) and Shulha and Cousins (1997).

2.3 The use of evaluation results

Lack of use of evaluation findings and recommendations has been a concern for evaluation managers for some time. Evaluation planning needs to pay greater attention to the final use and users of evaluation results, and plan accordingly.²

Here are three areas to consider in relation to evaluation use and applying the DAC criteria.

- 1** How much and what kinds of information do potential users need? Should there be equal focus on each of the criteria, or will some information be more useful? For example, some recent evaluations of humanitarian action have focused on internally displaced persons (IDPs), and findings related to coverage have been particularly important in these evaluations.
- 2** When will the information be useful? For example, information on effectiveness and efficiency may be more useful in ongoing interventions.
- 3** 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?

² For further information, see Annexe 1 below, World Bank (2004) and Henry (2003).

section three

Working with the DAC criteria

3 Working with the DAC criteria

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

- a definition;
- an explanation of the definition;
- issues to consider in use of the criteria, including areas that can be problematic;
- key messages; and
- examples of good practice, taken from a cross-section of countries, agencies, sectors and natural disasters and complex emergencies (further details of these case studies, including in many cases full reports, can be found in the ALNAP Evaluative Reports Database, at www.alnap.org).

Table 1 provides a summary of the DAC criteria, for easy reference. The following issues should be included for consideration when using the DAC criteria.

Using the criteria in combination. The DAC criteria are designed to promote comprehensive evaluation of humanitarian action. For this reason, the criteria are complementary. For example, evaluation of effectiveness may show that objectives were met, but this does not necessarily mean that the objectives were appropriate for the entire affected population, were met efficiently, are sustainable, or feed into impact. 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 help to ensure that an evaluation covers all areas of the intervention. There is often some overlap

between criteria, and evaluators may find that the same data can be included under different DAC criteria. Deciding where to place findings is less important than ensuring that a comprehensive evaluation is carried out, using all of the criteria where relevant.

Using the criteria to determine key factors for success. Use of the DAC criteria to date has tended to focus on results rather than processes: for example, how many people were fed, rather than the reasons why food aid was successful or not. But there is nothing inherent in the criteria which stops evaluators using them from asking ‘why’ questions – why was the project/programme effective, why was it efficient or not, and so on. Using the criteria in this way will promote lesson-learning. So evaluators should ask both what happened, and why it happened.

When to use the criteria. Evaluation managers and evaluators need to determine whether using all of the criteria is relevant. Because they were formulated for evaluation of projects or programmes, some of the DAC criteria may be less relevant for policy-based or institutional evaluations. And, for example, ‘coherence’ may be less relevant for evaluating single- agency or single-project interventions, such as a discrete water project.

Are data available? To evaluate interventions against the DAC criteria, it is preferable to have measurable objectives and data for baseline and ‘results’ measures, but these are often missing or partial. If data are not adequate, they may have to be constructed from available evidence, for example by interviews with key stakeholders, or by using oral-history techniques. As noted in Section 2.3 above, the key is to note constraints to evaluation practice transparently, so that the reader is able to assess the credibility of the methods employed, and results drawn from them.³

Cross-cutting themes. There are eight cross-cutting themes which should be considered when using the DAC criteria: local context; human resources; protection; participation of primary stakeholders;⁴

³ On oral history techniques, see Thompson (2000). On participatory techniques in humanitarian action, see ALNAP (2003).

⁴ Primary stakeholders, also known as beneficiaries, are the intended recipients of support and aid.

coping strategies and resilience; gender equality; HIV/AIDS; and the environment. Attention to these themes is included under the discussion of the individual criteria below. Not every evaluation needs to include every cross-cutting theme, but if any are to be left out there should be a clear rationale for doing so. For further details of the cross-cutting themes, see Annexe 2.

The DAC criteria and the logframe. The logical framework (or logframe) is currently the most common planning tool in development assistance, and its use is also widespread in humanitarian action. Project/programme planners therefore need to be able to link logframes to use of the DAC criteria, in order to plan how the intervention for which they are responsible will be monitored and evaluated. The main areas usually included in the logframe are activities, inputs, outputs and outcomes – also known as the results chain.⁵ The links between the DAC criteria and the results chain are discussed under each criterion below.

⁵ For definitions of these terms, see the OECD-DAC (2002) *Glossary of Key Terms in Evaluation and Results Based Management* (<http://www.oecd.org/dataoecd/29/24/2754804.pdf>).

Table 1 Summary definitions of the DAC criteria

Criterion	Definition	Main use*
Relevance/ Appropriateness	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.	<i>All evaluation types except those with a mainly institutional focus.</i>
Connectedness	Connectedness refers to the need to ensure that activities of a short-term emergency nature are carried out in a context that takes longer-term and interconnected problems into account.	<i>Evaluations assessing institutional structures and partnerships.</i>

Table 1 Summary definitions of the DAC criteria *continued*

Criterion	Definition	Main use*
Coherence	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.	<i>Joint evaluations, large-scale evaluations and those with a focus on policy.</i>
Coverage	The need to reach major population groups facing life-threatening suffering wherever they are.	<i>All evaluation types except those with a mainly institutional focus.</i>
Efficiency	Efficiency measures the outputs – qualitative and quantitative – achieved as a result of inputs. This generally requires comparing alternative approaches to achieving an output, to see whether the most efficient approach has been used.	<i>All evaluation types where adequate financial information is available.</i>
Effectiveness	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 criterion of effectiveness is timeliness.	<i>Single-sector or single-agency evaluations.</i>
Impact	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 intended and unintended, positive and negative, macro (sector) and micro (household).	<i>Multi-sector, multi-agency evaluations; joint evaluations; sector-wide evaluations.</i>

*All of the criteria will be useful in most evaluations to some extent. This column selects evaluations types where each criterion will be particularly useful. For further details on the types of evaluations noted in the third column, see ALNAP Training Module 4 (at <http://www.alnap.org/training.html>).

3.1 Relevance/appropriateness⁶

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 that can be used at different levels. 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 most appropriate activity for achieving this; alternatives could be food or cash for work, or measures to improve the functioning of local markets (OECD-DAC, 1999). 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.

Most of the cross-cutting themes in Annexe 2 can be considered under this criterion, for example whether the project/programme promoted participation, whether gender analysis was carried out, and whether women's and men's coping capacities were supported or undermined. In relation to the logframe results chain, relevance can be used to evaluate the wider elements of the intervention such as the overall goal or outcome, while appropriateness can be used to evaluate inputs and activities.

⁶ All definitions in this section are taken or adapted from OECD-DAC (1999), unless otherwise noted. Many of these definitions draw on Minear (1994).

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 takes into account local context. Interventions are more likely to be relevant and appropriate if they: (i) are begun with an adequate needs assessment; and (ii) show understanding of and support for 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 (women, men, girls and boys, different social groups), 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, given seclusion norms.

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? (See, for example, reports of experience in Kosovo (ALNAP, 2001), and the IFRC example of good practice 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.

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 conduct a relevant and appropriate response (the IFRC example of good practice below includes further details).

Key messages

- Relevance and appropriateness are complementary criteria used to evaluate both the wider goal of an 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 example: assessment of relevance/appropriateness 1

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

Background

Part of the evaluation of the response by the International Federation of the Red Cross and Red Crescent (IFRC) to the 1999 earthquake in Turkey was a thorough analysis of the relevance and appropriateness of the response in terms of: (i) 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.

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 expectations among the affected population, 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 urbanisation at a policy level, including 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:

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 (p 83).

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) 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 of 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 example: assessment of relevance/appropriateness 2

WFP evaluation of food aid for relief and recovery in Somalia

Background

This evaluation assesses three-year support to some 1.3 million people, with 63,000 million tonnes 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: 'contribute to a broader framework for integrated rehabilitation programmes in Somalia, while maintaining flexibility to both grasp development opportunities and respond to emergency situations' (p 4). 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. Arguments in favour of food aid included that: the country is usually in food deficit, populations in many locations are isolated from customary markets, and are doubly disadvantaged due to loss of primary occupations and assets. On the other side, 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 is also linked to questions of connectedness. For example, the evaluation of support to social institutions notes that:

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 (p 15).

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.

3.2 Connectedness

Definition

'**Connectedness** refers to the need to ensure 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, recovery 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. In complex emergencies where there is limited development activity, or in natural disasters where primary stakeholders are semi-permanently caught in the relief phase, sustainability may be difficult to achieve, which is why the focus in

EHA should often be on connections between humanitarian action, recovery and development.

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 government departments and/or development agencies, and adequate availability of funding post-response. Evaluation of connectedness is often linked to evaluation of impact, as both tend to consider longer-term consequences. In terms of the results chain in the logframe, evaluation of connectedness will mainly cover the linkages between outputs and outcomes. The assumptions column may also include information about risks related to connectedness.

Issues to consider

Evaluators should be sensitive to **relative expenditure on relief and recovery**. Substantial amounts of humanitarian aid are regularly spent on recovery, rather than relief (ALNAP, 2002). Recovery activities should be evaluated against appropriate indicators, which may not be the same indicators used to evaluate relief (eg people 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, for example 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**. Development of and support to capacity is central to ensuring that the effects of interventions are not lost. 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 example of

good practice below, 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, civil society and other partners at various levels is built by the intervention. ALNAP's research (2004) has shown that local capacity and the quality of local staffing is one of the key factors in success or failure of interventions.

Key messages

- Large-scale relief programmes can have significant long-term consequences, for example on local power structures, government capacity, gender equality or the environment.
- Evaluators should determine whether funds have been used for relief or recovery, 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 or local capacity is supported and developed.

Good practice example: 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. Linking relief, rehabilitation and development (LRRD), or ensuring connectedness, is usually a central theme in evaluations of ECHO-funded interventions. The manual states: '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, 1999, p 12). 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. Concerning one intervention, the evaluation notes:

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 (ECHO, 2001a, p 13).

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).

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 concerning the overlapping nature of relief and rehabilitation work. It notes that: '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 (ECHO, 2001b, p 10).' 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, 2004).

Sources: ECHO (2004a) *Evaluation of ECHO's Global Plan 2000-Angola. Evaluation Report 305*. Brussels: ECHO; ECHO (2004b) *Central America. Hurricane Mitch Global Plans 1998 and 1999. Health Sector-2. Evaluation Report 300*. Brussels: ECHO.

Good practice example: assessment of connectedness 2

Evaluation of DEC India earthquake appeal funds

Background

The Disasters Emergency Committee (DEC) evaluation of British NGO interventions after the earthquake in Gujarat in January 2004 is unusually detailed, resulting from collaboration between eight specialists from three organisations located in India and the UK, and based on three visits by the UK team members between March and October 2004. The evaluation used the Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief 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 how connectedness needs to be viewed from macro and micro perspectives, from the level of national institutions to that of local livelihoods. The evaluation examines the key area of timing, noting that few agencies made a strategic review at the crucial point when turning from relief to recovery. It analyses also 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 better utilised their

resources by pressuring the various levels of government to direct resources to the most vulnerable groups, as the political context meant that NGO support more likely replicated rather than replaced what the government would have done anyway. 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/resilience support is a key one that is often missed in EHA. This evaluation considers the trade-offs in water harvesting between speed and quality, that is between whether to build as many water-control structures as possible or to ensure the greatest community mobilisation. 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.

3.3 Coherence⁷

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

Assessment of coherence should focus on the extent to which policies of different actors were complementary or contradictory. This can involve any type of policy – for example, on promoting gender equality, participation or environmental protection.

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 has become increasingly important given recent close links between humanitarian and military actors in Kosovo, Afghanistan and Iraq. 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

⁷ Some evaluation team members piloting this guide thought that coherence should be dropped as a separate criterion, and included with coordination. Others thought it was too demanding to ask most evaluators to assess humanitarian response against this criterion, and that a new definition should be developed based on using the policies of the agency or agencies being evaluated as a standard against which to evaluate. The constraints to assessing coherence have been integrated into this section.

promoting the return of refugees to their host country, while another is opposed to such an approach (OECD-DAC, 1999). Another example is coherence between government import policies and food-aid distributions after the 1998 floods in Bangladesh, to stabilise the price of rice (Beck, 2005). 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 proved 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.⁸ The evaluation of coherence should complement that of coordination, with coherence focusing on the policy level, and evaluation of coordination focusing more on operational issues. For more on coordination, see under effectiveness in Section 3.6 below.

As far as the logframe is concerned, coherence is mainly addressed in terms of outcomes, that is the intermediate or longer-term effects of contributions of different actors to humanitarian action.

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 civilian–military relations. They also need to be able to live with the consequences of evaluating coherence, being one of the more ‘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

⁸ For more information on coherence, see Macrae and Leader (2000).

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.

Recently, military forces have been increasingly involved in humanitarian assistance, in some cases supporting humanitarian agencies and 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.⁹ 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. From this perspective, *lack* of coherence may be preferable. Because of the increased involvement of military forces, evaluators need to pay close attention to the mandates, agendas and principles of different actors, and evaluate whether these mandates contradict or complement each other.

Key messages

- Evaluation of coherence may 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.
- Key questions to ask in the case of coherence are: why was coherence lacking or present; what were the particular political factors that led to coherence or its lack; and should there be coherence at all?

⁹ For further details see IASC (2004) ([http://www.reliefweb.int/rw/lib.nsf/db900SID/DPAL-62GCWL/\\$FILE/ocha-civmil-28jun.pdf?OpenElement](http://www.reliefweb.int/rw/lib.nsf/db900SID/DPAL-62GCWL/$FILE/ocha-civmil-28jun.pdf?OpenElement)) and Chapter 4 of ALNAP (2004).

Good practice example: assessment of coherence¹⁰

Evaluation of UNHCR's emergency preparedness and response, Kosovo

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 number 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, and may have been negative for the well-being of many people affected by the crisis:

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' (ALNAP, 2001, p 72).

Given NATO's significant role in both military and humanitarian action (eg 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 that might have produced greater attention to coherence.

¹⁰ Only one example of good practice on the evaluation of coherence could be identified from ALNAP's Evaluative Reports Database.

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, and 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 unauthorised 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 are outlined from the perspective of both partners.

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. This illustrates again that coherence may not always be preferable or produce positive results. The evaluation also discusses 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.

3.4 Coverage

Definition

'The need to reach major population groups facing life-threatening risk wherever they are.'⁴⁴

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?' The Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief is clear concerning who should receive support (Box 3). Coverage is linked closely to effectiveness, as objective statements in humanitarian action, which are assessed with reference to effectiveness, often refer to numbers or percentages of the population to be covered.

Evaluation of coverage usually takes place at three levels

- 1** international level, determining whether the resources provided in one emergency are adequate in comparison to other emergencies – evaluations could spend more time considering this issue – and a good example is DEC (2000) which examines this issue in relation to the Kosovo conflict;
- 2** national or regional level, determining whether support was provided according to need in different areas, and why or why not; and

⁴⁴ Evaluators piloting this guide had differing opinions on whether coverage should be a separate criterion, or included within 'effectiveness'. Following OECD-DAC (1999), it has been included here as a separate criterion.

3 local level (village, slum, community and/or refugee camp), determining who received support and why; information at this level should be broken down by social categories such as socioeconomic grouping, gender, age and ethnicity.

At the regional and local levels, 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, socioeconomic 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, socioeconomic grouping and ethnicity).

Evaluating whether protection needs have been met is a key element in the assessment of coverage. Evaluators should ask whether those who needed protection have received it. 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. An example is given below in the DANIDA good practice case.¹²

Box 3 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.

Source: <http://www.ifrc.org/publicat/conduct/code.asp>

¹² 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 that the current tools may not be adequate for this approach (ODI, 2003).

Issues to consider

Proportionality. Providing aid according to need is central to humanitarian action (see Box 3). EHA should consider whether aid was provided according to need at the three levels noted above: internationally, nationally and locally. This is a central feature of the use of the coverage criterion.

Differing perspectives on what constitutes need. What constitutes 'need', and therefore who is covered by humanitarian action, is often culturally determined. Evaluations have found that there are differing opinions on whether the whole of an affected population, or the most vulnerable within that population, should be covered. A good example is food aid, which agencies tend to target towards the most 'vulnerable', but which communities often share according to local cultural norms (ALNAP, 2004; 2002). Evaluators need to be sensitive to this issue and determine whether targeting practices, often determined by donor governments, are appropriate from the perspective of primary stakeholders. A practical approach is to look at an organisation's own standards for coverage, and evaluate how far these have been met, and their relevance.

The evaluation report should present an estimate of the **proportion of those in need covered**, expressed as a percentage, rather than an absolute number. Evaluations 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 persons (IDPs) may require 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 has been recently published by SIDA.⁴³

In the logframe, coverage is usually incorporated in results statements and indicators relating to the numbers and types of the

⁴³ A multi-agency 10 country IDP review was undertaken in 2003/5; for details see <http://www.sida.se/sida/jsp/sida.jsp?d=418&a=3435&searchWords=internally%20displaced%20persons>.

affected people targeted. Results statements should be clear concerning the numbers to be covered, as well as particular groups being targeted. Terms such as 'vulnerable groups' should be disaggregated by socioeconomic status and sex.

Key messages

- Evaluating proportionality, that is whether aid has 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.
- Equity questions are central to analysis of coverage. Coverage should consider equity through both geographical analysis and a breakdown of data by relevant socioeconomic categories, such as gender, socioeconomic grouping, ethnicity, age and ability.

Good practice example: 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 concentrates 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 analysing a number of years is well placed to assess. It usefully contrasts interventions that attempt

to deal with short-term needs (eg 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 that 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 example: assessment of coverage 2

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

Background

The purpose of this evaluation 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 comprised four expatriates (one WFP staff member, 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 lack of data typical in such evaluations: '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' (p 17).

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:

- 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;
- impact of security concerns and restrictions on travel which have made access to some geographical areas problematic, eg in parts of Burundi;
- ways in which the PRRO has modified its approach away from generalised relief in favour of a more targeted focus in some parts of the region;
- targeting between different income groups within refugee camps;
- targeting of refugees outside camps, a population that may be missed in evaluations;
- the checklist on meeting WFP commitments related to gender equality 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.

The report notes also the problematic cultural issue of individual targeting: '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' (p20).

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 Efficiency

Definition

'**Efficiency** measures the outputs – qualitative and quantitative – achieved as a result of inputs. This generally requires comparing alternative approaches to achieving an output, to see whether the most efficient approach has been used.'

Explanation of definition

Efficiency measures how economically inputs (usually financial, human, technical and material resources) 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 an 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. Efficiency links to other criteria including appropriateness of choice of intervention, and effectiveness (see the two good practice examples below).

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, and these factors often determine why an intervention was efficient or not.

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 logframe. In some logframe, financial information is included at the input level, and this will provide a direct lead into the evaluation of efficiency.¹⁴

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 uses of resources.

¹⁴ For further information including downloadable resources, go to <http://www.mango.org.uk/guide/section.aspx?ID=5>.

- 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. This will ensure an appropriately balanced focus on both lesson-learning and accountability.
- Response 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.

Good practice example: assessment of efficiency 1

Evaluation of DEC 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 DEC agencies, in order to report to the UK public on how and where its 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, including 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, ie 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 (this area could also have been considered under appropriateness);
- 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 relies on cost implications of general practice followed, such as warehousing practices and transportation costs.

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 (2004) *Independent Evaluation of DEC Mozambique Floods Appeal Funds: March 2000 – December 2000*. London: DEC.

Good practice example: assessment 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. As is the case in 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 institutional, socioeconomic and environmental issues, linking these three areas together. The method involved document review, interviews with key staff, and beneficiary interviews. The discussion of efficiency is linked to effectiveness and impact of the intervention concerning changes in the local environment.

Evaluation of efficiency

The evaluation includes a detailed discussion of the costs of the firewood project in relation to other possible uses of funds (pp 99–110). Firstly the evaluation notes how it was hampered by lack of accurate financial data: ‘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’ (p 100). It is important that evaluations make such constraints clear. The evaluation goes on to examine key areas in relation to efficiency:

- 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 options for reduction of gender-based violence, 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);
- 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 efficient options that are also likely to reduce the extent of rape and other sexual and gender-based violence.

Source: UNHCR (2004) *Evaluation of the Dadaab Firewood Project, Kenya*. Geneva: UNHCR.

3.6 Effectiveness

Definition

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 criterion of effectiveness is timeliness.'

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 adolescent girls and the disabled.'

Assessing effectiveness in this case involves:

- examining the main reasons why 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 with a 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 – although this is quite rare in EHA – and at the very least there should be a plausible argument made concerning causality.

In the logframe, effectiveness is covered in the contributions of outputs to achieving outcomes. Evaluation of effectiveness is therefore linked to evaluation of impact and longer-term effects of the intervention.

Issues to consider

Use and benefit of resources. Where possible, evaluations need to go beyond assessing activities and begin to examine who uses and benefits from resources provided, with data disaggregated by sex, socioeconomic grouping and ethnicity. This may be a difficult area to evaluate in ongoing operations, or where baseline or monitoring data are incomplete. Even in these cases however, primary stakeholder interviews can reveal important information on use and benefits.

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, examining cross-cutting themes will be a useful support – for example, 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 as the number of houses to be built or amount of food to be distributed,

rather than the effect of the building or distribution. Evaluators should point this out where it occurs. The evaluator's role is to attempt to evaluate the intervention against stated objectives or agency or government standards, 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.

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 provision of support, goods and services achieved, according to the perceptions of key stakeholders? The DFID/WFP example of good practice below provides information on both the timing of the intervention as a whole, and distribution schedules for food aid.

Key messages

- Effectiveness measures objectives that are established in agency planning documents, and which can be considered as intermediate between outputs and outcomes.
- When analysing 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 analysing 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.

Good practice example: 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 to 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.

Evaluation of effectiveness

The evaluation included both quantitative and qualitative methods. Quantitative data were collected in 2,644 randomly selected households in villages throughout the project zone. Qualitative data were 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 logframe;
- assessment of criteria used for selection of beneficiaries, including primary stakeholder views of these criteria (an area which can also be assessed under coverage);

- analysis of implementation mechanisms, including levels of community participation;
- targeting accuracy, disaggregated by sex and socioeconomic grouping (again, this could be considered under coverage);
- examination of the resources provided – both the size of the ration and the commodity mix – including the reasons why they were provided (this could also be assessed under the relevance/appropriateness criterion);
- adequacy of distribution schedules; and
- the affected population’s view of the intervention.

Source: DFID (2004) *Emergency Food Aid to Flood-Affected People in South-western Bangladesh: Evaluation report*. London: DFID.

Good practice example: assessment of effectiveness 2

WFP evaluation of food assistance and support for repatriation of Iraqi and Afghan refugees in Iran

Background

WFP had 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 members. The evaluation team visited 7 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 these have been met. For example, the first objective is to 'Ensure the basic food needs for survival of the refugees', one corresponding output 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 percentages 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 consideration of the political environment and planning processes.

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.

3.6.1 Coordination¹⁵

While not a 'formal' DAC criterion, coordination is an important consideration in the evaluation of humanitarian action. Coordination cuts across several criteria, but is included here under the heading of effectiveness, as the two are closely related. Coordination has been defined as:

'the systematic use of policy instruments to deliver humanitarian assistance in a cohesive and effective manner. Such instruments include strategic planning, gathering data and managing information, mobilising resources and ensuring accountability, orchestrating a functional division of labour, negotiating and maintaining a serviceable framework with host political authorities and providing leadership' (Minear et al, 1992).

Whereas coherence focuses on whether policies of different actors are in line with each other, coordination focuses more on the practical effects of actions of governments and agencies – for example, whether they join cluster groups, whether they discuss geographical targeting, and the extent to which information is shared.

¹⁵ This section draws on Reindrop and Wiles (2004) and Van Brabant (1999). Some evaluators piloting this guide thought that coordination should be a separate criterion but there was no consensus on this.

Questions to consider during the evaluation of coordination:

- 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 channelled in a coordinated fashion, or individually by donors to suit their own strategic aims?
- What were the main constraints and supports to coordination? How was good coordination achieved, and is it replicable in other situations?
- Did coordination lead to improved effectiveness and impact, for example an increase in humanitarian space? How did this come about?

Three key issues to consider concerning coordination:

1 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). But numerous evaluations point to coordination as a problematic area that will pay dividends if improved (ALNAP, 2002), and hence it should be included in evaluations and lesson-learning exercises where possible.

2 The roles of the host government and other local institutions. For example, were objectives and activities 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).

3 Non-traditional partners. A further issue concerns how coordination with non-traditional partners such as the military is organised. It is important to solicit viewpoints of the different stakeholders involved on this type of coordination.

3.7 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 intended and unintended, positive and negative, macro (sector) and micro (household).'

Explanation of definition

Whereas assessment of effectiveness examines whether intermediate objectives have been achieved and fed into outcomes, assessment of impact usually examines the longer-term consequences of achieving or not achieving those objectives, and the issue of wider socioeconomic 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 the wider socioeconomic and political contexts – as can be seen in the good practice example below, from Rwanda. Because of their longer-term focus, evaluation of impact and connectedness are often closely linked.

Issues to consider

Is evaluating impact relevant in all cases? Because of its wider scope, assessment of impact may not be relevant for all evaluations, particularly those carried out during or immediately after an intervention. Changes in socioeconomic 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 that have not often been deployed in evaluations of humanitarian action to date. Therefore, evaluation of impact should be attempted only where: a longitudinal approach is being taken; there are data

available to support longer-term analysis; the evaluation team includes specialists in socioeconomic 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 whether changes that have taken place are the result of the intervention or of some other factor, such as other interventions, or socioeconomic or political forces. To help to deal with this issue, ‘informal’ control groups of the affected population who have not received assistance can be interviewed, as setting up formal control groups through quasi-experimental design is usually not feasible in EHA.

Phrasing of results statements, and the logframe. Impact can be considered a higher-order measure and thus relates to the outcome column in the logframe, as well as to the intervention goal. Goal statements in the logframe are often vaguely worded with few quantitative targets stated. Evaluators may therefore need to refer to agency policy to determine how the intervention fits with the longer-term goals of the agency.

Evaluation of impact needs to cover cross-cutting themes such as support to livelihoods, human rights, and gender equality. It is well suited to evaluation of these themes because they often involve longer-term and wider socioeconomic processes.

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?
- Assessment of impact may be the most challenging aspect of carrying out an evaluation – the evaluator has to determine the timeframe to be considered, and ensure that there is a plausible assumption that the intervention contributed to impact achieved. When including impact 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.

Good practice example: assessment of impact 1

Joint evaluation of emergency assistance to Rwanda

Background

To date, the Joint Evaluation of Emergency Assistance to Rwanda (JEEAR) is the largest and most comprehensive evaluation of humanitarian action, involving 52 consultants and researchers. It also sets standards for the joint assessment of the 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, despite 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 socioeconomic functions; this necessarily leads to an analysis of political structures that largely determine humanitarian response and impact.

Evaluation of impact

In the Rwanda case, the lack of intervention is considered in two parts: (i) analysis of historical factors which explained the genocide; and (ii) detailing 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 that used in evaluations of other crises: conflict and its aftermath in Kosovo, the effects of Hurricane Mitch, and interventions in Afghanistan. In each of these cases, decisions were made to carry out single-agency and single-sector evaluations, which largely missed the political nature of the event and the response to it. In the Kosovo and Afghanistan cases, this led to a lack of attention by evaluators to issues of protection and human

rights (ALNAP, 2004; 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 geopolitical significance – rather than listing events 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 at what might have happened given other circumstances and different kinds of intervention.

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

Good practice example: assessment 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 reports each being written by one of these consultants. The purpose of the evaluation was both to 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 of 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 immunisation. Building on the discussion of these basic statistics, it goes on to consider the wider impact of ECHO's intervention on the healthcare system and public services. It shows that even with humanitarian action's short timeframe 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 have 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;
- that 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
- that dependency is being created, with NGOs dominating the direction of healthcare and government counterparts at district level having limited involvement in priority-setting and planning.

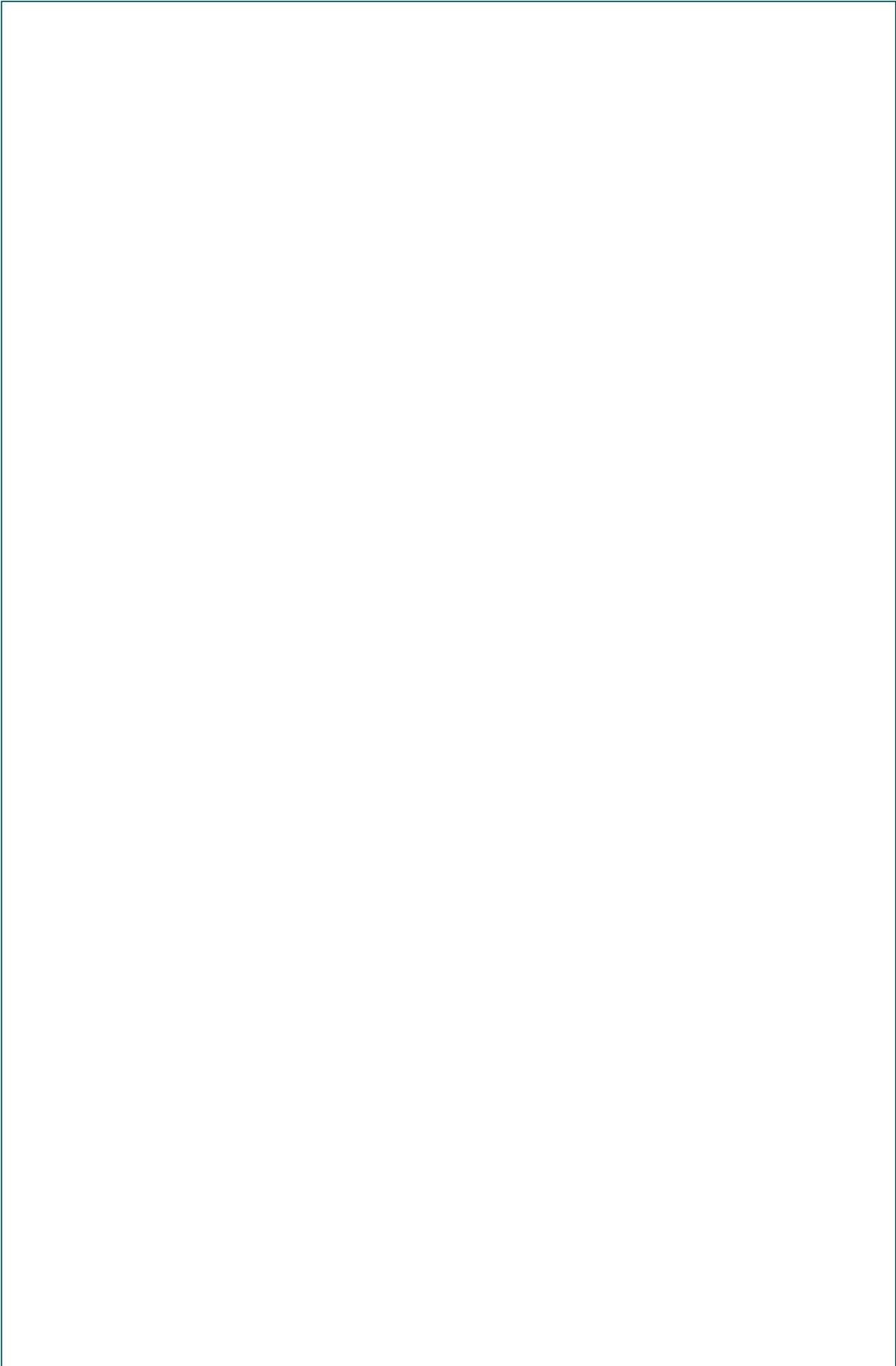
Reasons for achieving impact are also discussed in some detail, for example the evaluation notes that:

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' (p 14).

The report also notes the importance of community involvement in the achievement of impact.

Source: ECHO (2004) *Sierra Leone, ECHO Global Plan 2000/Intervention Plan 2001. Health, Nutrition, Water and Sanitation – 2001*. Brussels: ECHO.



section four

Methods for the evaluation of humanitarian action: pointers for good practice

4 Methods for the evaluation of humanitarian action: pointers for good practice

EHA differs in some ways from general evaluation and from the 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.

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 are available online (at www.alnap.org). Wood, Apthorpe and Borton (2001) and Hallam (1998) are further useful sources.

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 questioned is the method! Evaluations should in particular note: the reasons for the 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 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. First, 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. Second, EHA should be as comprehensive as possible; if the views of the main group 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 women and men, to primary stakeholders from different socioeconomic groups and ethnicities, and to girls and boys. 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 re-live their past experiences 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, socioeconomic 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.

Ensure a focus on social process and causality. EHA tends to concentrate on what happened rather than why it happened, which does not maximise support to lesson-learning. Evaluators need to build questions into their evaluation questionnaires such as: 'What were the reasons why 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 to any endeavour (for example, a desire to empower primary stakeholders is one kind of bias, and 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 them, usually helps to reduce evaluator bias.

Last but not least: **integrate the DAC criteria**, or provide the rationale for not using the criteria.

references

ALNAP (2005) *Protection – An ALNAP Guide for Humanitarian Agencies.* London: Overseas Development Institute (http://www.alnap.org/pubs/pdfs/alnap_protection_guide.pdf).

ALNAP (2004) *Review of Humanitarian Action in 2003. Field Level Learning.* London: Overseas Development Institute.

ALNAP (2003) *Humanitarian Action: Improving Monitoring to Enhance Accountability and Learning.* ALNAP Annual Review 2003. London: ALNAP/ODI.

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.

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.

Beck, T (2005) *Fighting Poverty while Supporting Recovery from Major Natural Disasters. Bangladesh Case Study.* Washington DC: World Bank and the ProVention Consortium.

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.

Hallam, A (1998) *Evaluating Humanitarian Assistance Programmes in Complex Emergencies.* London: ODI, Good Practice Review 7.

Henry, G (2003) 'Influential Evaluations' in *American Journal of Evaluation* Vol 24, No 4.

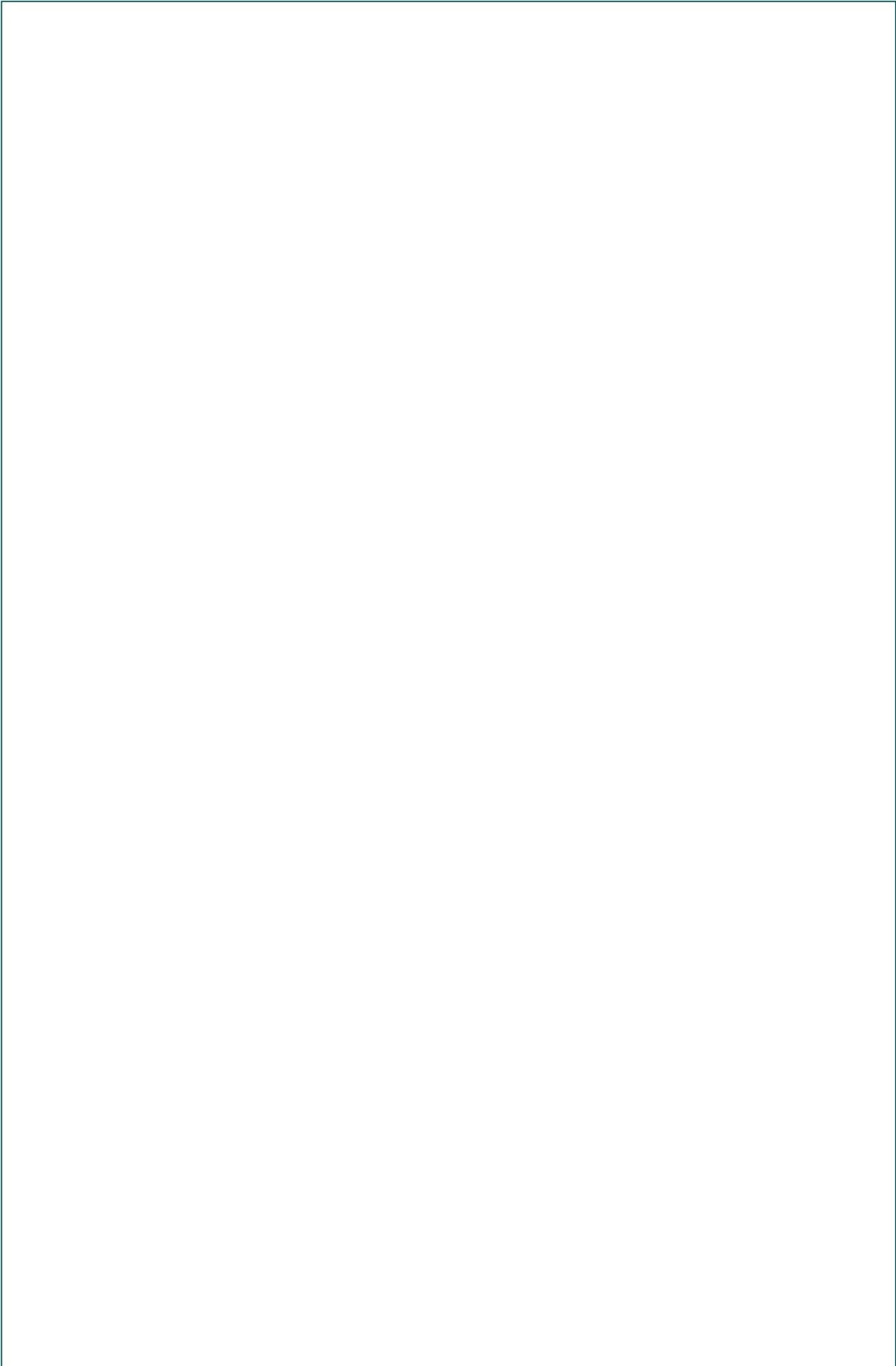
House, E (2004) 'The Role of the Evaluator in a Political World' in *Canadian Journal of Program Evaluation* Vol 19, No 2, pp 1–16.

IASC (2004) *Civil-Military Relationship in Complex Emergencies.* Geneva: Inter-Agency Standing Committee.

Lackey, Jill et al (1997) 'By Whose Standards? Reflections on Empowerment Evaluation and Grassroots Groups' in *Evaluation Practice* Vol 18, No 2, pp 137–146.

Longley, C and D Maxwell (2003) *Livelihoods, Chronic Conflict and Humanitarian Response: A Synthesis of Current Practice.* London: ODI, Working Paper 182 (http://www.odi.org.uk/publications/working_papers/wp182.pdf).

- Macrae J and N Leader (2000)** *Shifting Sands. The Search for 'Coherence' between Political and Humanitarian Responses to Complex Emergencies*. London: ODI, HPG Report 8 (<http://www.odi.org.uk/hpg/papers/hpgreport8.pdf>).
- Miner, L (1994)** 'The International Relief System: A Critical Review', paper presented to the Parallel National Intelligence Estimate on Global Humanitarian Emergencies, Meridian International Centre, Washington DC, 22 September.
- Miner, L, U Chelliah, J Crisp, J Mackinlay and T Weiss (1992)** UN Coordination of the International Humanitarian Response to the Gulf Crisis 1990–1992. Providence: Thomas J Watson Institute for International Studies, Occasional Paper 13.
- ODI (2003)** *According to Need? Needs Assessment and Decision-Making in the Humanitarian Sector*. 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 (1998)** *Review of the DAC Principles for Evaluation of Development Assistance*. Paris: OECD.
- OECD-DAC (1991)** *Principles for Evaluation of Development Assistance*. Paris: OECD
- 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' in *Evaluation Practice* Vol 18, No 3, pp 195–208.
- Thompson, P (2000)** *Voice of the Past: Oral History*. Oxford: Oxford University Press.
- UNICEF (2001)** *Managing Monitoring and Evaluation Activities. Monitoring and Evaluation in Crisis and Unstable Contexts*. New York: UNICEF.
- Van Brabant, K (1999)** *Opening the Black Box. An Outline of a Framework to Understand, Promote and Evaluate Humanitarian Coordination*. London: Overseas Development Institute.
- Wood, A, R Apthorpe and J Borton (eds) (2001)** *Evaluating International Humanitarian Action. Reflections from Practitioners*. London: Zed Books.
- World Bank (2004)** *Influential Evaluations: Evaluations that Improved Performance and Impact of Development Programs*. Washington DC: World Bank.



annexe one

Joint Committee on Standards for Educational Evaluation

Joint Committee on Standards for Educational Evaluation

To improve performance, EHA needs to be contextualised within wider good practice in the field of evaluation in general. In 1994, the Joint Committee on Standards for Educational Evaluation (JCSEE) 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 (<http://www.wmich.edu/evalctr/jc>). As these standards are widely used in the evaluation community, it makes sense to employ them to guide understanding of ways of working with the DAC criteria.⁴⁶

The central points of relevance to EHA of the four basic evaluation standards – utility, propriety, feasibility, and accuracy – are summarised below after relevant components of the standards are noted. Not all of the standards are quoted below, and more details can be found at: <http://www.eval.org/publications.asp#EDUCATIONAL%20EVALUATION%20STANDARDS>.

⁴⁶ DAC (OECD-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. For a recent discussion of international standards, see *New Directions for Evaluation*, Winter 2004. 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/Publications/GuidingPrinciples.asp>. All of these publications have been taken into account in developing this guide.

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 judgements 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 promoting evaluation use, investigating how and why results from previous evaluations have been used in commissioning agencies. Recommendation-tracking matrixes can also be employed.

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 programme 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, regardless of 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.

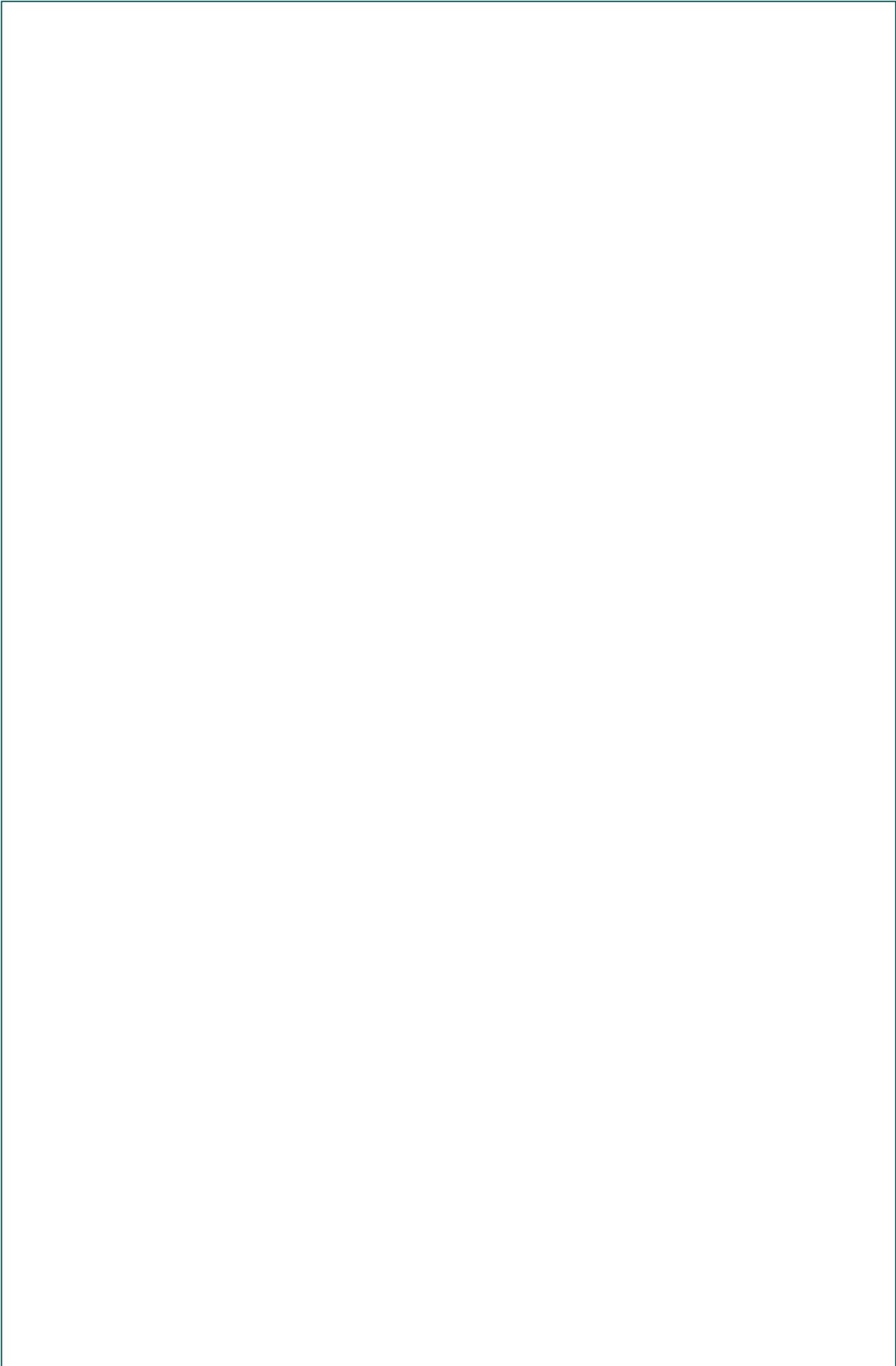
Relevance to EHA. Evaluators need to make themselves aware of interest groups in the commissioning agency and of what their interest in the evaluation is likely to be. Any potentially controversial conclusions and recommendations need to be justified as clearly as possible.

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 programme being evaluated.

- **Programme documentation.** The programme being evaluated should be described and documented clearly and accurately, so that the programme is clearly identified.
- **Context analysis.** The context in which the programme exists should be examined in enough detail, so that its likely influences on the programme 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 programme 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, and to describing and justifying methods used and sources and analysis of information gathered (ALNAP, 2002).



annexe two

Checklist of themes to be covered in EHA

Checklist of themes to be covered in EHA

This checklist is offered as a resource for evaluators to cross-check that they are covering key elements when using the DAC criteria. It is not a comprehensive list but contains several important EHA themes. For suggestions on where cross-cutting themes might apply to the DAC criteria, see Section 3 above, under individual criteria.

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 in the results of a particular project or programme.

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 turnover; field–HQ relations and communication; the role of national staff; and training and learning practices. On how field workers learn, see the ALNAP Review of Humanitarian Action 2003 (<http://www.alnap.org/RHA2003/index.htm>).

Protection. Humanitarian protection has emerged as a key issue in humanitarian response over the last few years. However, evaluation of humanitarian protection remains limited. ALNAP defines humanitarian protection as: ‘the challenge of making states and individuals meet their humanitarian responsibilities to protect people in war, and filling-in for them as much as possible when they do not’ (ALNAP, 2005, p 21). For further information, see Protection – An ALNAP Guide for Humanitarian Agencies (2005) (http://www.alnap.org/pubs/pdfs/alnap_protection_guide.pdf).

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 draft ALNAP handbook (at http://www.alnap.org/pubs/pdfs/GS_Handbook.pdf). Evaluators and evaluation offices are also accountable to primary stakeholders, in terms of both consultation and ensuring that evaluation results are used to benefit primary stakeholders to the maximum extent possible.

Coping strategies and resilience. 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. For further details on coping strategies and resilience in relation to humanitarian action, see Longley and Maxwell (2003), http://www.odi.org.uk/publications/working_papers/wp182.pdf.

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. On gender equality and humanitarian action, see CIDA's 2004 guide (at [http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Africa/\\$file/Guide-Gender.pdf](http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Africa/$file/Guide-Gender.pdf)).

HIV/AIDS. The spread of HIV/AIDS is becoming increasingly important to humanitarian action. The interactions between food insecurity and HIV/AIDS should be assessed in evaluation of countries or regions where HIV/AIDS is widespread. On evaluation of HIV/AIDS-related interventions, see the the bibliography in http://data.unaids.org/publications/irc-pub06/jc1126-constrcoreindic-ungass_en.pdf.

The environment. Evaluations should assess whether interventions have supported environmental sustainability, or whether they have harmed the environment.