EVALUATION AND AID EFFECTIVENESS

1

Guidance for Evaluating Humanitarian Assistance In Complex Emergencies

DEVELOPMENT ASSISTANCE COMMITTEE

OECD

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

FOREWORD

Evaluation is a key tool in efforts to improve accountability and performance in the operation of the international humanitarian assistance system. Historically, humanitarian assistance has been subjected to less rigorous and extensive monitoring and evaluation procedures than development aid. As the share of ODA allocated to humanitarian assistance has risen, and awareness of its complexity has increased, so the need to develop appropriate methodologies for its evaluation has become steadily more apparent.

Evaluations of humanitarian assistance operations which have been undertaken in recent years have varied enormously in terms of their approach, the available human and material resources, and management structures. The widely varying quality, scope and depth of these evaluations may be likened to a "methodological anarchy". As a consequence it has been difficult to judge the relevance and accuracy of individual studies, and the comparative analysis necessary to inform wider institutional learning has been severely hampered. Increasing the consistency and quality of evaluation methodologies would enhance the accountability function of evaluation. It would also contribute to institutionalising lessons learned and to identifying better methods of monitoring performance of humanitarian assistance operations.

This Guidance is aimed at those involved in the commissioning, design and management of evaluations of humanitarian assistance programmes principally within donor organisations but is also likely to be of use to UN agencies, NGOs and other organisations involved in the provision of humanitarian assistance. It is not intended as an exhaustive guide as specialised texts are available, but to complement the existing DAC Principles on Aid Evaluation by highlighting those areas which require special attention, the nature of the activities undertaken and the multi-actor, highly interconnected system by which the international community provides humanitarian assistance.

This work was commissioned by the OECD/DAC Working Party on Aid Evaluation. Overseas Development Institute, London (ODI) was contracted to prepare the draft guidelines, and the work has proceeded with the financial support of Danida, Finnida, the Netherlands Development Co-operation and the UK Department for International Development.

I would like to extend my appreciation of the extensive contributions received -- both in terms of financial support and substantive inputs -- which facilitated the completion of this publication.

Mr Niels Dabelstein
Chair, Working Party on Aid Evaluation

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ALNAP	Active Learning Network on Accountability and Performance in Humanitarian Assistance
DAC	Development Assistance Committee
DANIDA	Danish International Development Assistance
ECOWAS	Economic Community of West African States
ECHO	European Commission Humanitarian Office
FINNIDA	Finnish International Development Agency
JEFF	Joint Evaluation Follow-up Monitoring and Facilitation Network to the Rwanda Study
ICRC	International Committee for the Red Cross
IDP	Internally Displaced Persons
IFRC	International Federation for the Red Cross
NGO	Non-governmental Organisation
OAU	Organisation of African Unity
OECD	Organisation for Economic Co-operation and Development
TOR	Terms of Reference
UK ODA	United Kingdom Overseas Development Assistance
UN	United Nations
UNHCR	United Nations High Commission for Refugees
UNOSOM	United Nations Operations in Somalia
WFP	World Food Programme

GUIDANCE FOR EVALUATING HUMANITARIAN ASSISTANCE IN COMPLEX EMERGENCIES

1. Introduction

Due to the comparatively recent and limited body of evaluative literature on humanitarian assistance, identifying good practice has involved the use of a variety of complementary sources and methods and extensive consultations and iterations.

Altogether the process has involved:

- the review of at least 70 readily accessible humanitarian assistance evaluation reports and synthesis studies;
- the circulation of a questionnaire to key individuals involved in humanitarian aid evaluations and the analysis of their responses;
- interviews with 65 individuals in 30 humanitarian organisations (funding, channelling and implementing agencies) in Paris, Rome, Geneva, New York, Washington and the UK.

Preliminary findings and/or early draft reports were presented at a number of fora involving evaluation managers and those with operational experience with humanitarian organisations. The most important of these fora was a meeting convened by Danida under the auspices of the DAC Working Party on Aid Evaluation in Copenhagen in January 1998 which involved representatives of 16 bilateral and multilateral donor organisations, 6 UN agencies, the IFRC and the ICRC and 5 NGOs. Subsequent drafts have been discussed at the regular meetings of the Working Party on Aid Evaluation and the final document was approved in July 1999.

As the difficulty of addressing different audiences within a single document became apparent, it was decided to target this guidance principally at Evaluation Managers within donor organisations and to produce a separate, more discussive review of good practice aimed at a broader audience including agency field personnel. The two documents have been prepared in tandem; an earlier version of this guidance was used as the basis for the more general review which was recently published in the Good Practice Review series of the Relief and Rehabilitation Network (RRN Good Practice Review No. 7 'Evaluating Humanitarian Assistance Programmes in Complex Emergencies' by Alistair Hallam. September 1998, ODI, London).

2. Complex Emergencies, Humanitarian Assistance Programmes and their Evaluation

2.1 Complex Emergency: Definition and Origins

The term 'complex emergency' was coined in Mozambique in the latter half of the 1980s. An important factor influencing its coinage and usage was the need for international aid agencies to acknowledge that the 'emergency aid' or humanitarian assistance needs were being generated by armed conflict as well as by

periodic 'natural disaster' events, such as cyclones and droughts, whilst avoiding mentioning use of terms such as 'war', 'civil war' and 'conflict' which were sensitive terms in the Mozambican context at the time. Since then, particularly following the end of the Cold War period, the international community has been more directly involved in efforts to provide humanitarian assistance in areas of ongoing armed conflict. As a consequence the term 'complex emergency' has entered widespread usage as a way of differentiating those situations where armed conflict and political instability are the principal causes of humanitarian needs from those where natural hazards are the principal cause of such needs.

The term is potentially confusing in that it implies that natural disasters cannot be 'complex' (and are somehow 'simple'), and similarly that conflict-related emergencies occurring prior to the 1980s (such as that in Biafra in 1968-71) were not 'complex' even though many of the characteristics of that emergency and the dilemmas faced by donor organisations and humanitarian agencies were very similar to more recent emergencies occurring in Eastern Europe, Asia and other parts of Africa. However, the term is useful in that does highlight the fact that situations involving political instability and armed conflict are often particularly complex contexts in which those involved in the provision of humanitarian assistance have to operate. It is the very complex context of such emergencies and the international community's system for responding to them, that requires the adoption of particular approaches and techniques by those involved in the evaluations of humanitarian assistance activities in response to complex emergencies.

2.2 Characteristics of Complex Emergencies and the International System for Responding to Them

It is a common saying that no two emergencies are the same -- each has unique origins and characteristics and it is potentially misleading to think in terms of a 'typical' complex emergency. Nevertheless, for the purposes of this document it might be helpful to highlight differences between 'typical' complex emergencies, other emergency contexts and the 'typical' development assistance context for which conventional aid evaluation approaches have been developed. Key characteristics of complex emergencies are:

⇨ *Intra-State rather than Inter-State Conflict*

The majority of conflicts are now intra-state (often termed civil wars) rather than inter-state. Most intra-state conflicts result from a disaffection by part of the population with the existing structures of governance and authority. In many situations the declared objective of opposition or 'rebel' groups is either to overthrow the government in power or to secede part of the country to establish a new autonomous entity. The degree of recognition or status given to opposition groups or warring factions by international organisations and the international community is often a highly sensitive issue. This context has important implications for the role of international agencies, the legal basis for their operations and the legal rights of affected populations within the affected countries.

⇨ *Difficulty in Differentiating Combatants and Civilians*

In many recent intra-state conflicts it has proven difficult to distinguish between civilians and combatants. Uniforms are often not worn. Combatants may be fed by their families or food procured from the local population either on a voluntary basis or through coercion. In many cases the intermingling of combatants and civilians is a deliberate policy. In such a situation humanitarian agencies are often unable to prevent assistance distributed for use by genuine civilians and vulnerable populations being used by combatants and warring factions.

⇨ ***Violence Directed Towards Civilians and Civil Structures***

In many contemporary conflicts civilians are deliberately attacked and their way of life undermined in order to displace particular social or ethnic groups or in some cases to actually eliminate them. As well as the more usual attacks on economic infrastructure (e.g. government buildings, factories, roads and railways) this may involve deliberate attacks on communities with the objective of instilling such fear in neighbouring communities and groups of similar social or ethnic background that they will seek refuge elsewhere. In this way territory, farmland and housing may be released for subsequent use by the faction responsible for the atrocity. Attacks may also be carried out on targets which play a special role in the cultural identity of particular groups (e.g. places of religious worship and culturally significant buildings).

⇨ ***Fluidity of the Situation on the Ground***

Whilst rapid-onset natural disasters (e.g. earthquake, floods, cyclones) often involve sudden and traumatic events the duration of the events is rarely more than a few days, though it may take months or years for the affected population to resume their pre-disaster livelihoods. Slow-onset natural disasters such as droughts are of at least several months duration and prolonged droughts may last for 2 years or more. However, their effects may take weeks and months to develop. Complex emergencies are often chronic situations lasting several years. In the case of Eritrea for instance the conflict lasted three decades and in the cases of Afghanistan, Mozambique and Angola almost two decades. Within such chronic contexts the situation can be highly fluid in particular geographical areas at particular times. Fighting may produce civilian as well as military casualties and threatened populations may often flee creating, more or less localised, displacement crises. Chronic problems are therefore frequently interspersed with situations requiring urgent responses by humanitarian agencies.

⇨ ***Lack or Absence of Normal Accountability Mechanisms***

Whilst the degree of functionality and freedom enjoyed by the press and judiciary may be limited in many states not affected by instability or conflict, in most complex emergencies the degree of functionality and freedom is either severely constrained or has been eliminated. Those involved in the conflict and those involved in trying to provide humanitarian assistance therefore operate in a context of absent or severely weakened national accountability mechanisms.

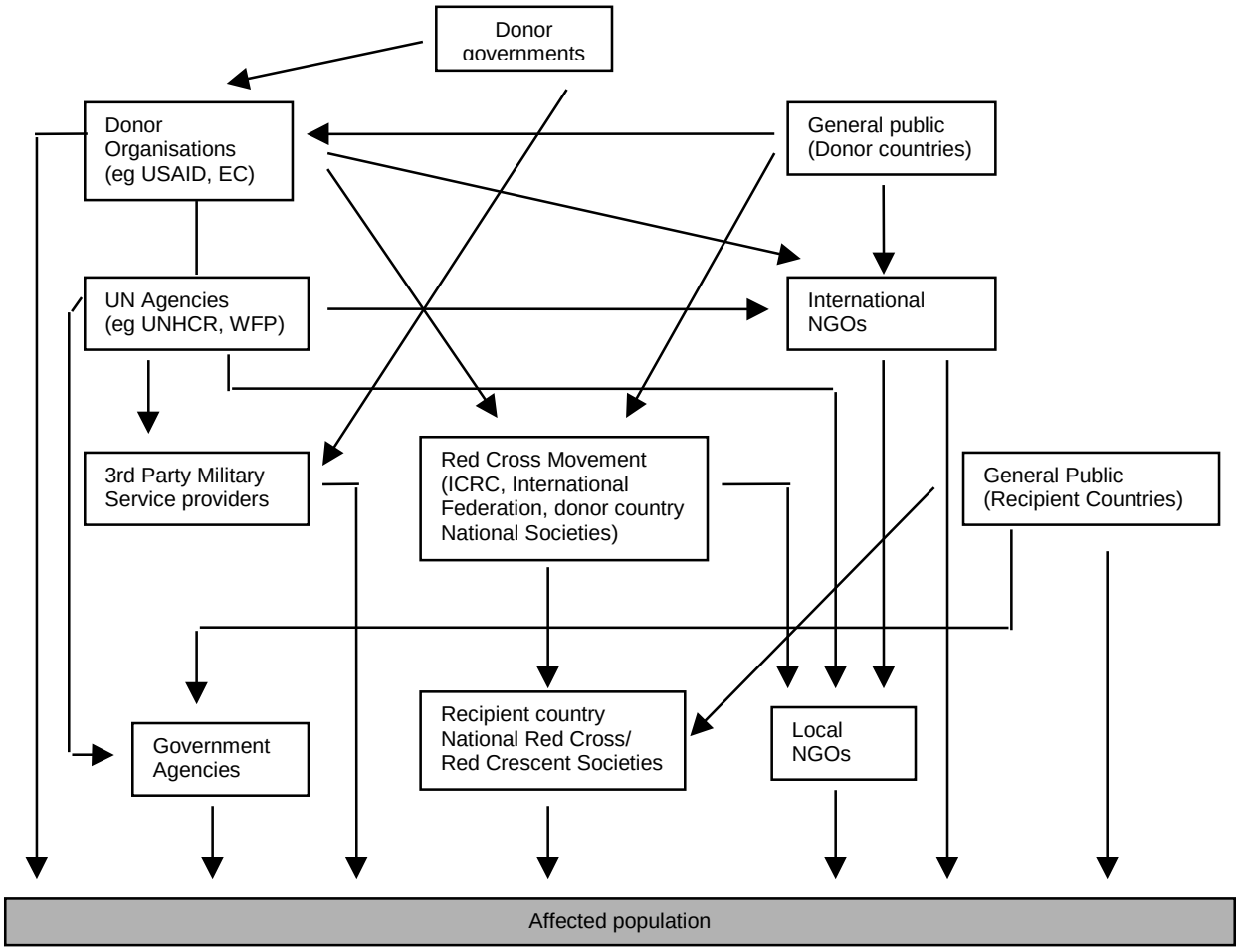
⇨ ***The Potential and Actual Development of War Economies***

The absence of accountable authorities in parts of or all of a conflict affected country often results in the development of economic activities which would normally be classified as illegal or semi-legal. Such activities may involve the exploitation and exportation of mineral deposits and natural resources, the cultivation and exportation of narcotics and the laundering of ‘dirty’ money and trading in arms. Often these activities are controlled by, or heavily taxed by, leaders of the warring factions who use the resources either for personal gain or to prosecute the conflict. Such ‘war economies’ enable the conflict to be prolonged and may develop to the extent where they actually provide an incentive to the conflict and its prolongation. Because of the international character of the transactions such illegal and semi-legal activity invariably involves the participation of commercial organisations based outside the affected countries in neighbouring countries, industrialised countries or other countries. In some cases such activity is carried out with the knowledge of the authorities.

⇨ ***The Potential for Humanitarian Assistance to Prolong the Conflict***

As a resource being provided into areas of ongoing conflicts, humanitarian assistance is potentially capable of being 'diverted' from the intended beneficiaries and controlled and taxed in the same way as 'war economy' activities. Reliable empirical evidence of the degree to which humanitarian assistance is diverted is often lacking even though such knowledge may be fundamental to understanding the impact of the assistance provided either on the intended target group or in potentially providing warring factions with additional resources.

Figure 1. Resource Flows Within the International Relief System



⇨ ***A Multiplicity of Actors***

The complexity of the international system for responding to complex emergencies cannot be overstressed. The preceding Figure (Figure 1) conveys some of the complexity by showing the principal routes of resource flows within the system.

The organisations commonly involved in the provision of humanitarian assistance include national relief structures (in areas where they are still operating) or relief structures associated with particular factions, local NGOs, UN agencies, international NGOs, the ICRC, the IFRC and the national Red Cross and Red Crescent Society. In some instances large bilateral and multilateral donor organisations may establish local

field teams with the objective of funding locally prepared projects and co-ordinating their overall activities. Diplomatic activity, by neighbouring states, powerful states and others seeking to achieve either a cessation of the conflict or prevent its escalation, is a common feature of many complex emergencies. Human rights agencies and monitors and organisations seeking to resolve the conflicts are increasingly active in areas of ongoing conflicts. In those situations where international peacekeeping or peace enforcement forces are deployed either by the UN or by regional bodies (e.g. NATO, OAU, ECOWAS, CIS) humanitarian activities will run in parallel to, or in concert with, the peacekeeping operations. Such military interventions may involve troop contingents from a variety of countries. In situations where displaced populations cross international borders, refugee agencies will become involved together with the governments of the asylum countries. It is not uncommon for neighbouring states to be involved (either overtly or covertly) in intra-state conflicts in neighbouring countries. States with traditionally strong links with the conflict-affected country or which perceive their strategic interests to be at stake may also play a role in relation to the conflict. A typical complex emergency may therefore involve the following organisations: 7-8 UN agencies; the Red Cross Movement (ICRC, IFRC and the National Society); 50 or more international and local NGOs involved in the provision of humanitarian assistance, human rights activities and conflict-resolution activities; military contingents operating either under a UN or a regional organisation mandate; and agencies controlled by or associated with the warring factions. The range of security and humanitarian activities may be funded by 20 or more official donors supplemented by funds raised privately. In addition a number of neighbouring and powerful states may be taking a keen interest in the course of the conflict and possibly seeking to influence the outcome by overt or covert means.

Given that central authority structures are generally weak or absent, co-ordination of such a multiplicity of actors becomes a major challenge - one that has not yet been fully met by the international humanitarian system. Co-ordination structures and mechanisms which exist within the UN system and the NGO community are rarely strong and effective. In addition, the wide range of actors involved in the response or with an interest in the outcome of the conflict almost guarantees that their goals will not be shared and may even be at odds with each other (see Box 1).

Box 1: An Example of Differences in the Goals of Key Actors

In managing refugee settlements near the Kenyan/Somali border in 1991, UNHCR had to balance a number of diverse interests and pressures. The camps initially provided a safe haven for those fleeing armed conflict, but they also attracted refugees from Somalia who were not displaced by war, as well as Kenyan nationals wishing to access repatriation benefits. Similarly, the goal of the Government of Kenya - rapid repatriation of the refugee population - was not shared by UN agencies. UNOSOM, for example, was advising against repatriation because of the continuing insecurity in large areas of Somalia (Kirkby, Kliest & O'Keefe, 1997:15, cited in Apthorpe and Nevile, 1998).

2.3 Differences Between the Evaluation of Humanitarian Assistance Programmes and 'Conventional' Aid Programmes

In terms of 'good practice', there are many ways in which the evaluation of humanitarian assistance programmes in complex emergencies is little different to the evaluation of aid programmes in general: Terms of Reference need to be prepared, teams selected, field study undertaken and reports and recommendations prepared. However, the very characteristics of complex emergencies and the way in which international assistance is organised and provided requires the explicit consideration of, and in many instances an emphasis upon, approaches to evaluation which are not typical of those used in relation to development projects and programmes. This section highlights the principal differences.

There is enormous variation in the scope of humanitarian assistance evaluations ranging from evaluations of a single project undertaken by a single agency at the bottom, to 'system-wide' evaluations of the international community's response to a major crisis at the top. Figure 2 indicates the principal levels in this range.

Figure 2. A Hierarchy of Evaluations

System-Wide Evaluation	Evaluation of the response by the whole system to a particular disaster event or complex emergency
Partial system Evaluation	Evaluation of a part of the system such as a thematic or sectoral study
Single Agency Response Evaluation	Evaluation of the overall response to a particular disaster event or complex emergency by a particular agency (funding agency, channelling agency, implementing agency)
Single Agency, Single Project Evaluation	Evaluation of a single project undertaken by a single agency

Within each level evaluations may also vary in terms of the relative emphasis given to lesson-learning as opposed to accountability.

⇨ *Working in Areas of Conflict and Instability*

The fact that many humanitarian assistance operations take place in a context of conflict and instability has several significant implications for the nature of the evaluation process.

Humanitarian agencies operate within a 'humanitarian space' that may be constrained to a greater or lesser extent by deliberately restricted access (roadblocks, attacks on aid convoys and personnel) as well as access being restricted through poor or damaged infrastructure and seasonal climatic factors. Evaluators need to make themselves fully aware of such constraints and make allowance for them in their assessment.

Humanitarian assistance programmes are not just about the provision of material and technical assistance. The population being assisted may well be subject to a range of human rights abuses such as attack, murder, rape, harassment and exclusion from accessing basic services by the armed forces and groups directly involved in the conflict or by others exploiting the breakdown of law and order. Invariably such abuses of human rights are committed by armed males and as a consequence there is a strong gender

dimension to the abuses; women and children are particularly vulnerable. Humanitarian assistance evaluations therefore need to assess the security situation in the area of operation and the occurrence of human rights abuses of the civilian population and consider the approach adopted by the agencies to that situation. Whilst humanitarian assistance evaluations to date have been patchy in their coverage of such issues, it is now widely recognised that future evaluations need to assess the 'humanitarian space', the security situation and the protection needs of the affected population. These are substantially different concerns from those of 'conventional' aid evaluations.

The nature of the subject matter means that evaluations of the humanitarian aid responses to complex emergencies are undertaken in areas that have recently experienced active conflict and may be continuing to experience instability and insecurity. Whilst it is rare for evaluation teams to be deployed to areas of active conflict, this is not unknown, and may require the provision of additional insurance cover for war zones to personnel not covered by schemes already available for the personnel of the agency commissioning the evaluation. In terms of how the context of conflict and instability affects the work of the evaluation team, however, probably the most important factors are the impact of the events upon those being interviewed and the extreme sensitivity of the subject matter.

Whoever the team is interviewing, it is important that, whilst retaining their objectivity, they bear in mind and attempt to empathise with the experiences that their interviewees have endured. This applies regardless of whether the interviewees are officials within the government or a 'rebel' faction, relief workers, military personnel within the factions and in peacekeeping contingents, or civilians within the affected population who received assistance provided. The affected populations will have just experienced a civil war during which they may well have been forcibly displaced, had relatives and friends killed, either in the fighting or during atrocities committed against civilians, and seen their personal, social and perhaps cultural identities shattered. Psycho-social trauma may affect much larger numbers than is evident to an outsider, particularly if unfamiliar with the local language and untrained in the diagnosis of post-traumatic stress disorders. Exposed to such extreme experiences and perhaps having themselves witnessed massacre sites, it is not unheard of for members of evaluation teams to also experience mild forms of traumatic stress disorders. (In the same way that it has become the practice of offering counselling to returning relief personnel, the practice of offering counselling to returning evaluators should not be ruled out).

Civil wars greatly exacerbate and deepen existing fault lines within a society and may well create new cleavages in previously coherent groups. For instance, tensions may develop between those who stayed through the conflict and those who sought safety outside the country; those who came to support a new faction and those who remained loyal to a former government or faction. Such is the intensity of feeling and the polarisation of affected societies that objectivity and the truth become difficult concepts to uphold. It is not uncommon for evaluators working in complex emergencies to experience two intelligent and articulate adults giving completely contradictory versions of the same event (see Box 2). In such situations evaluators may have to accept that their search for 'the truth' may never be successful. The implication of this for their work is that the veracity of information collected cannot be taken for granted; it will require constant checking and cross-checking referring to different sources using the technique of

Box 2: Examples of Different, and Intensely-Held, Perspectives

The Joint Evaluation of Emergency Assistance to Rwanda was presented to a three day, Government sponsored Conference in Kigali in September 1996. The finding in Study 3 that international assistance had been biased towards the refugee population and that 60% of all assistance- provided to the region had gone outside Rwanda was hotly disputed by Government officials. They claimed that their own analysis showed that the bias was much worse with 95% of the assistance having gone outside the country. The evaluators were themselves accused of being biased towards the refugees and against the new Government.

'triangulation'. When writing the report care will need to be taken to ensure the careful wording of particularly sensitive sections.

⇨ *Confronting the Lack of Vital Information*

A product of the characteristics of Complex Emergencies is that key information on a range of matters of vital significance to evaluators is often unavailable. Whilst evaluators of development assistance programmes are also often faced with a lack of information on key indicators or decisions, such problems are considerably and routinely more serious in the case of evaluations of humanitarian assistance programmes in complex emergencies. Some or all of the following factors may combine to produce a situation where information which is vital to evaluators either does not exist or is not easily accessible: the multiplicity of actors; the fluidity of the situation; the difficulties of working in the context of war and instability; the frequent absence of baseline data on the condition of the affected/target populations prior to the assistance intervention; the failure by some agencies to monitor key indicators; a lack of agreement on standardised monitoring procedures and protocols among agencies; and the difficulties of adhering to normal standards for recording discussions and decisions and maintaining filing systems. Thus, evaluators of humanitarian assistance programmes are routinely faced, not only with the need to compensate for missing project data, but also have to contend with a lack of information on the context, the precise sequence of events during the period and the goals and (often changing) policies of different actors at different stages of the emergency. In other words they have to construct 'histories' and 'pictures' of the vital information to serve as a form of baseline from which to judge the appropriateness and effectiveness of the policies and the projects.

'Total system' evaluations will nearly always require a substantial degree of 'baseline' construction, to fit alongside evaluations of response policies and selected illustrative projects. However, evaluations of single projects - even comparatively 'technical' projects - do not escape the need to construct such baselines. For instance the evaluation of a project by Agency 'B' to establish treatment centres for dysentery cases, or a project by Agency 'C' to establish water production and distribution systems for displaced populations, will require the following components: context setting; the explanation of the events in the project area; clarification of the needs that the projects were attempting to address; and the relationship between the projects undertaken and those undertaken in the same area by other agencies. Such explanation may reveal that alternative actions by the agencies or other actors earlier in the emergency may have prevented the dysentery outbreak or avoided the population being displaced. More effective action earlier by agencies 'B' and 'C' or by other agencies and actors may have avoided the need for the subsequent interventions. Where such alternative courses are revealed the use of cost-effectiveness analysis becomes particularly pertinent as the objectives of the interventions may have been achieved at lower cost or greater impact might have been achieved at the same cost.

⇨ *A Greater Requirement for Policy Evaluation*

In the past evaluations of humanitarian assistance tended to focus on projects and utilise conventional project evaluation techniques. However, thinking has shifted and it is now believed that humanitarian assistance evaluation requires a greater emphasis upon policy evaluation techniques than is often the case for 'conventional' aid evaluation. The reasons for this are several.

First, the fluidity of the context and the complexity and interrelatedness of the response system reduces (though by no means eliminates) the value and effectiveness of project evaluation techniques which require the separation of cause and effect. Explanation based on the separation of cause from effect is often not possible in complex systems composed of numerous interdependent relationships where the direction of influence may well be circular rather than linear. Thus methods which are more common in historical or philosophical research are often more productive than those traditionally employed in the social sciences.

Such methods acknowledge the complexity and interdependent nature of events in the real world and ask not, 'did *x* cause *y*?' but rather, 'what happened?' and 'why?' In other words, in order to understand and be able to deal with situations and structures, they seek to build narratives about specific events and processes, rather than theorising grandly and establishing causal relationships.

Second, humanitarian assistance is presently confronted by a range of major policy questions (such as whether it may prolong conflicts and how best to provide protection to civilians in ongoing conflicts) and evaluation has a key role to play in addressing such questions.

Section 3.2 discusses policy evaluation techniques and the appropriate mix between policy and project evaluation techniques.

3. Guidance for Evaluation Managers

3.1 'Upstream' Measures to Enhance the Effectiveness of the Evaluation Process

A theme of many recent humanitarian assistance evaluations has been that the effectiveness of the evaluation process has been limited, in many cases significantly, by the lack of:

i) clearly stated objectives for the overall programme and its various sectoral and project components, and

ii) monitoring information necessary for assessing the performance of projects and thus of the overall programme

⇨ ***Articulating and Clarifying Objectives***

Encouraging explicit and clear statements of objectives will require concerted effort over a period of time at several different levels of the humanitarian system:

Strategic Frameworks will need to be established which set the overall objectives for the international community's response (Strategic Frameworks are currently being developed and piloted by the UN system for Afghanistan and are planned for the Great Lakes region and Sierra Leone).

Country Strategies or Response Strategies will need to be articulated by donor organisations and other organisations involved in the response to a particular country or complex emergency.

Logical Framework Analysis is increasingly being used by donor organisations and some implementing agencies as a way of articulating the Goal, Purpose, Outputs and Indicators for humanitarian assistance projects. Some donor organisations have made the use of LogFrames mandatory for funding proposals above a specified funding level.

Use of the LogFrame does appear to bring a number of benefits, it can: significantly increase transparency in the setting of objectives; make more explicit the conceptual frameworks underlying interventions; help to avoid confusion between means and ends; and highlight the 'level' at which an organisation is intervening. However, concerns have been voiced that, for humanitarian assistance programmes, the log-frame may be used in an over-restrictive, potentially inflexible way and that while useful for looking at inputs and outputs, it has proved less useful in looking at overall policy goals, social ideals and higher principles. For the method to be used effectively it is important that the LogFrame for each project is

reassessed periodically and, if necessary, revised. A critical but often relatively neglected element of the method is the 'critical assumptions' column and it is vital that the assumptions made at the project design stage are periodically reassessed during the life of the project.

When considering undertaking an evaluation of humanitarian assistance provided in a complex emergency, Evaluation Managers should assess whether clearly articulated objectives are available not only for the individual projects funded but also the overall objectives of their own organisation and their partners the international humanitarian system generally. Where objectives are not clearly articulated the difficulties of undertaking an effective evaluation will be substantially increased. Whilst not proceeding with the study is always an option in such situations, the scale of the expenditure or the importance of the issues raised during the response may warrant proceeding with the evaluation. Allowance will need to be made for the expectations made of the evaluation and the fact that the evaluators may have to retrospectively construct the objectives for the different actors and interventions. Whilst the retrospective construction of LogFrames is far from ideal, it may be necessary to attempt this in relation to key projects for which LogFrames were not prepared at the time of the intervention. Where articulated objectives are available, evaluators should consider whether they were realistic and whether the level of resources deployed was appropriate to both the objectives and the needs that were being addressed.

⇨ ***Improving the Availability of Monitoring Information***

Current monitoring and reporting systems for humanitarian assistance programmes often do not take full account of the needs of ex-post evaluations and the effectiveness of the evaluation process is considerably hampered. Addressing this problem will require concerted action throughout the system to ensure that:

i) agencies improve their monitoring systems and use data collection systems that facilitate ex-post evaluations and cross-agency, cross-programme comparisons;

i) agreement is reached on the key indicators that should be monitored by all agencies. Potentially the Sphere Minimum Technical Standards process underway within the NGO community will provide such agreement on key indicators (see Box 9);

ii) a commitment by agencies to facilitate evaluations through the management of filing and information systems so that key reports showing the decision-making process are easily accessible.

Whilst (i) and (ii) are unlikely to be achieved in the short term, (iii) is more readily achievable.

Too often, evaluators start their work unaware of all the relevant internal documentation or useful related studies undertaken by other agencies in the same area or similar agencies in the same country. Sometimes, the problem is simply one of poor file management and valuable time can be spent searching for key documents. It is essential, for the organisation's institutional memory as well as for the evaluation process, that important information is retained and readily accessible. Evaluation Managers could seek to avoid such problems by insisting that agencies involved in the evaluation collate their files and information to be used by the evaluators well in advance of work starting. Key documents, such as Situation Reports, and Monthly Project Reports, should be collated for easy access by the evaluators. Country offices involved in the evaluation should be included in such efforts as they are often well placed to know of related studies by other agencies.

3.2 Deciding What to Evaluate, How and When

⇨ *What to Evaluate*

Potentially there are many options available to Evaluation Managers in deciding the scope, focus and objectives of an evaluation. Should only the response to the emergency be considered or should the actions that were, or might have been, taken prior to the emergency also be considered? If the emergency was prolonged over several years should the evaluation look back over the whole emergency or just consider the most recent or critical periods of the emergency?

Often there is a tendency for implementing agencies to evaluate only those projects they were involved in implementing and for donor organisations to evaluate only those projects which they supported through the provision of financial and material and logistical assistance. However many other possible approaches exist and should be considered. The highly inter-related nature of the different types of intervention produces strong arguments in favour of collaborative studies with other agencies that enable related interventions to be considered together. For instance, it is rarely possible to assess the impact of food aid on morbidity and mortality without also taking into account health and sanitation programmes (see Box 3). Similarly it rarely makes sense to consider the effects of a programme in one village without taking into account programmes in neighbouring areas, to which those villagers may have had access. Whilst collaborative studies are organisationally more demanding and require a degree of consensus and trust on the part of the agencies, efforts should be made at the conception stage to explore the possibility of developing a joint evaluation together with other agencies involved in the overall response or whose programmes shared a similar geographical or sectoral focus.

Box 3: How a Lack of Monitoring Information Can Limit the Effectiveness of Evaluations

In response to the 1991-92 drought in Mozambique, the UK ODA provided over 15 grants to NGOs for seeds and tool distributions. In examining narrative reports submitted to ODA the evaluation team found that only one of the agencies unambiguously indicated the dates when the seeds were distributed. It was, therefore, not possible for the team to state what proportion of the total tonnage of seeds had reached farmers before the onset of the rains. Only one of the agencies supported undertook surveys during the harvesting period to assess the production levels achieved and thus the overall effectiveness of the seeds and tools interventions. Thus two key indicators of effectiveness and impact were not, therefore, available for the majority of the agencies supported.

The dysentery epidemic which affected the Great Lakes Region during 1994 was responsible for more deaths than any other single non-violent cause. Mortality rates in the IDP camps in the Zone Turquoise appear to have been very high. However, many of the camps were not covered by epidemiological surveillance systems until after the epidemic had peaked. For instance surveillance in Kibeho, the largest and most problematic camp, only started 11 weeks after the area became secure. As a result, the precise dimensions and dynamics of the dysentery epidemic in this important camp of around 100,000 people will never be known. Given this lack of information it was not possible for evaluators to assess the effectiveness and impact of the nutritional, medical and public health interventions carried out by the agencies which worked in Kibeho and the other camps.
[Summarised from Clay et al. 1995 and Borton et al. 1996.]

So far the record of humanitarian assistance evaluations in assessing 'humanitarian space', the security situation, and the extent to which the protection needs of the affected population have been met, has been very patchy (see Box 4). In large part this has stemmed from such issues not being regarded as being a major concern of evaluation and the work of evaluators. Evaluations of humanitarian assistance programmes should include an assessment of the humanitarian space available, the level of human rights abuses and the measures taken to provide protection to the affected population. It will be important to explore the gender dimension of the abuses and the extent to which the agencies have taken account of this in the design of their programmes.

Even where an agency may have chosen not to directly address protection needs because they were the responsibility of another organisation (such as a peacekeeping force or a human rights monitoring team), the evaluators should still assess the adequacy of the protection mechanisms provided by other organisations and the extent to which the agency being evaluated was correct in its assumption that 'it was someone else's responsibility' and whether it might have done more to enhance protection. Such actions might include lobbying for international action, providing witness to atrocities or informing local authorities of their obligations under the Geneva Conventions.

In some contexts the protection needs of the affected population might be directly related to the design and conduct of the humanitarian assistance programme. For example, aid agencies may inadvertently lead armed opponents to refugees in hiding or the actual receipt of aid may make the beneficiaries more vulnerable to harassment or killing (see Box 5).

Box 4. Humanitarian 'Space'

Humanitarian agencies operate within a 'humanitarian space' that may be constrained to a greater or lesser extent by deliberately restricted access (roadblocks, attacks on aid convoys and personnel) as well as access being restricted through poor or damaged infrastructure and seasonal climatic factors.

Humanitarian space is a dynamic concept. Levels of access and availability of resources can change regularly during an aid operation. Agencies can themselves influence the humanitarian space available to them. Successful negotiation, for example, may open new routes through contested areas. Agencies can also reduce the space available to them by sticking rigidly to their mandate, even though flexibility might prove more effective. Agencies may sometimes take a principled stand and refuse to supply relief inputs where an unacceptably high proportion of these are being diverted by combatants to fuel the war. Humanitarian space may thus be restricted in the short-term, in the hope that this will lead to more freedom to operate effectively over the longer-term.
[Adapted from Hallam 1998]

It is now widely recognised that as an injection of resources and commodities into a conflict, humanitarian assistance has the potential to fuel local war economies and even in some contexts may actually prolong the conflict. It is therefore necessary for evaluations of humanitarian assistance programmes in complex emergencies to consider the existence and nature of the local war economies and the extent to which the humanitarian assistance provided may have fed into these economies through diversion or taxation. Important factors that should be considered are: the overall scale of the war economy and the relative scale of contribution to it by diverted or taxed humanitarian assistance; and steps taken by the agencies to avoid or minimise the level of diversion or taxation through the selection of particular delivery channels, supervised distributions and end-use monitoring.

Complex emergencies frequently result in the displacement of populations, either as refugees into neighbouring countries or as internally displaced persons (IDPs) within the affected country. The assistance needs of these populations are invariably substantial and the focus of attention by humanitarian agencies. However, experience has repeatedly shown that the host communities in the areas where the displaced populations settle or are accommodated are also affected by virtue of hosting the incomers but that their assistance needs are often inadequately addressed by the international community. The impacts of displaced populations on host communities are not necessarily all negative but certainly there are numerous negative impacts such as: the development of camps on farmland and the destruction

Box 5. Food Delivery and Protection in Liberia

In Liberia in 1987 a WFP food delivery in a contested area was followed by a massacre of part of the population who had received the food. An investigation by a human rights organisation concluded that the massacre was intended to demonstrate to the population that, although they might be able to receive aid from the international community, they were still under the control of the armed group concerned.

'Report on the Sinje Massacre' Justice and Peace Commission, Monrovia, Liberia October 1996.

of crops and fuel-wood sources; the increased incidence of disease, insecurity; and the overloading of local markets and services resulting in localised inflation and declines in the service levels previously enjoyed by the host community. Such negative impacts and the tendency for humanitarian assistance to be focussed on the displaced populations may result in resentment by the host community towards the displaced populations with potentially serious implications for humanitarian operations, such as border closures and the need to transfer the displaced to other areas. Evaluations of assistance to displaced populations should therefore also assess the needs of the host communities and the extent to which these were addressed in a timely manner.

Complex emergencies invariably have a variety of regional dimensions including the hosting of refugees by neighbouring countries, the use of transport routes through neighbouring countries by humanitarian agencies and possibly the support (overt or covert) of neighbouring states for one or more of the groups or factions involved in the conflict. In more than one case an inability to access the affected population in the conflict-affected country combined with a large refugee outflow has resulted in assistance efforts that were subsequently viewed as having been disproportionately focussed upon the refugees rather than the population which remained in the conflict-affected country. Evaluations need to be aware of such regional dimensions and perhaps explicitly factor their consideration into the scope of the study – with visits to the capital cities of neighbouring countries as well as to the conflict-affected country as well as to the refugee camps, and the explicit consideration of the 'balance' of the response between the conflict-affected country and the neighbouring countries in terms of the overall response and the agency being evaluated.

⇨ ***How to Evaluate***

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

As noted in Section 2.3, evaluations of humanitarian assistance usually focus on operations (i.e. projects) rather than on policy issues. Donor organisations are generally more comfortable evaluating the projects through which a policy was implemented, rather than evaluating the assumptions which lie behind a particular policy. However, several recent studies have explicitly considered policy issues (e.g. Apthorpe et al. 1996; Karim et al. 1996) and there are strong arguments for the more direct consideration of policy issues in the evaluation process.

Policy evaluation involves the consideration of the rationale and objective set for the programme and the extent to which they were able to predict and explain the context and the outcomes. It therefore examines the framework of understanding, beliefs and assumptions that make individual projects possible as well as desirable. Policy evaluations might therefore focus on the mandates of agencies and their ideologies and institutions, often in a framework which compares one agency, or set of agencies, with another. Policy evaluations seek out the inherent tensions or contradictions in policy objectives, through tools such as discourse analysis and logic-of-argument analysis (see Box 6). For example, a donor may seek to combine relief, development and peace objectives in its policy towards complex emergencies. Discursive, philosophical, theoretical analysis may show these to be counteractive or contradictory. At the same time, if empirical evidence demonstrates that actual operations have been successful, despite convincing theoretical argument against a policy being justifiable, then policy evaluation should take this into account and the policy be modified accordingly.

Policy evaluation involves a process of 'validating' through argument, rather than 'verifying' through some 'scientific' process, the various interpretations. In this respect, it is a 'discipline comparable with judicial procedures of legal interpretation, having a logic of uncertainty and of qualitative probability'

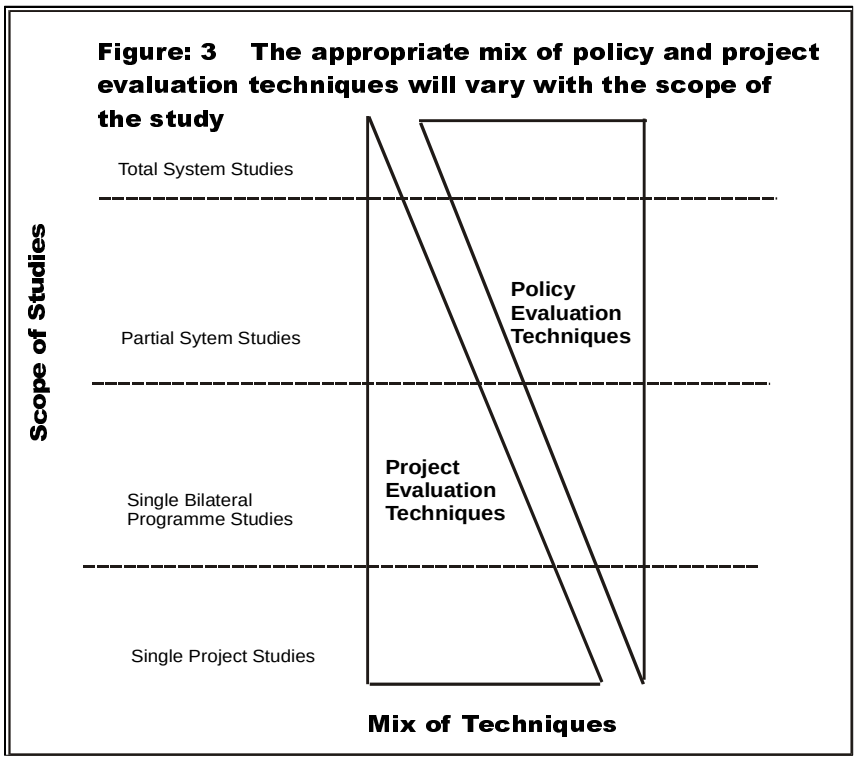
(Ricoeur, 1979:90 in Apthorpe and Gasper). Being concerned with the sum that is qualitatively greater than the (project) parts, policy evaluation is generally less handicapped by the lack of quantitative data than project evaluation.

Box 6. Discourse analysis

Discourse analysis involves the recognition of the fact that 'there is a plurality of values and arguments available for thinking about any specific policy issue. Analysis, therefore, has to be part of a process in which these several points of view are taken into account or directly included in the analysis.' [White, 1994 in Gasper and Apthorpe, 1996]. Discourse analysis involves a high level of multi-disciplinary analysis, in comparison to more traditional research methods. Sociology and political theory are among the disciplines which have made extensive use of discourse analysis techniques.

Policy goals or objectives are seldom clear or sharp enough to serve as criteria against which to test management and performance. They are not immediately evaluable in their own terms. In particular, those involved in policy evaluation must take the stated objectives of policies, projects and evaluation ToR, and translate them into something evaluable. Recognising that goals and objectives may change in response to changing circumstances on the ground and changes in the level of resources available is another requirement for best evaluation practice. This is particularly true of complex emergencies, which are, by their nature, unstable, yet also very often protracted or recurring.

Figure: 3 The appropriate mix of policy and project evaluation techniques will vary with the scope of the study



What is an appropriate balance between the techniques of policy evaluation and project evaluation? All evaluations will generally involve a mix of policy and project evaluation techniques. However, the precise balance between the two will depend upon the scope of the evaluation. Total system studies will invariably involve use of policy evaluation techniques, though with the use of detailed project-level analysis to illustrate the points being made. At the same time, single project studies cannot simply ignore the context in which operations take place. In attempting to understand situations and structures, to analyse a particular set of events and processes, the construction of a narrative history is a powerful tool. A range of different actors may be asked to tell their story, recognising that what they say represents the truth as they see it, or the truth as they would like the evaluator to see it. In this way, a partial understanding of

someone else's view of reality may be developed. The stories of many different actors are then added to the available documentary evidence to construct the narrative history. The narrative history is therefore more than a simple chronology. It details not just what happened and when, but also who was involved, and why, and links significant events.

The narrative history may be used by the evaluator to help to make judgements about events and processes, to explain why actors did what they did and with what effects and draw practical conclusions accordingly. When many different stories are accumulated, consistent patterns may emerge which help to explain actions. Alternatively, a framework for understanding events may arise out of previous studies and be confirmed by the evidence revealed in the narrative history. The narrative history approach is particularly suited to evaluating complex emergencies because it focuses on qualitative as well as quantitative methods; because it allows the evaluator to focus on actions, processes and intentions; and because it highlights the competing agendas of the diverse range of actors. Furthermore, use of the narrative history approach does not preclude the use of more analytical models, such as logical frameworks or cost-effectiveness reviews, to examine specific components of an emergency assistance programme.

Where it is decided to directly evaluate policies, the question that then arises is whether the norms on which the policies were based should be evaluated? Ideally they should, as this will provide a richer understanding of the response and the policy issues involved as the underlying norms are not necessarily the ones under which people were actually operating.

Box 7. Pertinent Provisions in International Human Rights and Humanitarian Law

Universal Declaration of Human Rights 1948 especially Articles 3 and 5.

International Covenant on Civil and Political Rights 1966, especially Articles 6 and 7.

The four Geneva Conventions of 1949, especially common Article 3; Articles 23, 55 and 59 of the Fourth Geneva Convention; Articles 69 to 71 of Additional Protocol I of 1977; Article 18 of Additional Protocol II of 1977.

Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment 1984, especially Article 3.

International Covenant on Economic, Social and Cultural Rights 1966, especially Articles 1, 11 and 12.

Convention on the Rights of the Child 1989, especially Articles 6, 22, 24, 37 and 38.

Convention on the Status of Refugees 1951, especially Article 33 and the Protocol relating to the Status of Refugees 1966.

Convention on the Prevention and Punishment of the Crime of Genocide 1948

Convention on the Elimination of All Forms of Discrimination Against Women 1979

[Drawn from the Sphere Project (1998), 'Humanitarian Charter and Minimum Standards in Disaster Response – Chapter 1' Geneva, and Darcy, James 'Human Rights and International Legal Standards: What Do Relief Workers Need to Know' RRN Network Paper 19, February, ODI, London.]

The operation of the international humanitarian system has long been hampered by the lack of widely accepted benchmarks and a clear legal framework. The legal framework is composed of a patchwork of legal provisions drawn from different branches of international law, human rights law, humanitarian law (the Geneva Conventions) and refugee law and which are designed more to protect against threats to security (physical, economic, social, political) than to meet people's immediate needs (see Box 7). Inevitably the lack of widely accepted benchmarks and a clear legal framework has also hampered the work of evaluators.

In the past few years, however, there has been a considerable effort by the humanitarian community to strengthen the framework in which humanitarian assistance programmes take place. This effort has led to the development of the widely-accepted NGO/Red Cross Code of Conduct (see Box 8) as well as the development of technical standards through the Sphere Project (see Box 9), both of which should be of considerable value to evaluators.

Box 8. Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in disaster relief

1. The humanitarian imperative comes first
2. 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
3. Aid will not be used to further a particular political or religious standpoint
4. We shall endeavour not to act as instruments of government foreign policy
5. We shall respect culture and custom
6. We shall attempt to build disaster response on local capacities
7. Ways shall be found to involve programme beneficiaries in the management of relief aid
8. Relief aid must strive to reduce future vulnerabilities to disaster as well as meeting basic needs
9. We hold ourselves accountable to both those we seek to assist and those from whom we accept resources
10. In our information, publicity and advertising activities, we shall recognise disaster victims as dignified humans, not hopeless objects

[Headlines from the Code of Conduct published in 1994 by the IFRC, Geneva]

Evaluators should be conversant with recent writings on humanitarian policy and practice, and, in the evaluation report, reference should be made to key works. This can help locate the evaluation's findings in a widely discussed framework, and can give the evaluation greater credibility than would be achieved if the report is seen simply as a subjective view of one individual.

Clarity over the objectives of the evaluation will make the whole process of the evaluation clearer and easier. Evaluation Managers should allow adequate time for this critical stage of identifying the objectives of the study. Advice from a range of sources inside and outside the organisation should be drawn upon. Evaluation Managers with little previous experience of evaluating humanitarian assistance programmes ought to involve a humanitarian assistance evaluation specialist early on in the process or consult with other Evaluation Managers with more experience. However, the process of agreeing the objectives will have to be carefully managed to avoid a 'shopping list' of unprioritised and potentially conflicting objectives.

⇨ ***When to Evaluate***

When the evaluation is best carried out will depend in large part on the objectives of the study and the context of the operations to be studied. However, there are often strong arguments in favour of carrying out an evaluation of a humanitarian assistance operation or a complex emergency whilst it is still ongoing rather than await the winding down of the operation and the end of the emergency. The need for evaluators to construct a narrative and 'baseline' requires that they interview many of the key actors before they are re-deployed to operations in other parts of the world; the turnover of humanitarian agency personnel is notoriously high. The ability of the target population and beneficiaries of assistance to accurately recall events such as the timing and routing of their displacement, the timing

Box 9. The SPHERE Project

During 1997 and 1998 a coalition of European and North American NGOs involving over 640 personnel drawn from 228 organisations collaborated in the development of a Humanitarian Charter and Minimum Standards in the five key sectors of water and sanitation; food security; nutrition; health services; and shelter and site selection. The draft Charter and the Standards were published in December 1998 and the following 2 year phase will consist of process of testing, dissemination and institutionalisation. (http://www.ifrc.org/pubs/sphere)

of their first relief distributions and the identity of the providing agency is reduced with each week that passes after the peak of the operations. It should also be borne in mind that the peak of assistance operations often occurs weeks or in some cases months after the peak of the emergency. Ideally therefore the first field visits should take place within 6 months of the peak of the operations, though the elapsed times for many of the humanitarian evaluations undertaken to date has been between 12-18 months. Such early deployment will probably involve careful consideration of the security risks and insurance considerations, as well as greater attention to sensitive questioning where interviewees may have been through traumatic events. However, the quality of the evaluation is likely to be greater if it is not delayed significantly.

The complexity of the subject matter and the need to construct narratives and baselines calls for flexible and tailored approaches to the humanitarian assistance evaluations. One possible approach is to stagger the evaluation so as to enable a first phase or 'pre-study' to construct the narratives and baselines and identify the key issues on which to focus in the main phase of the study. Experience has shown that the main issues that emerge in humanitarian assistance evaluations are rarely apparent at the start of the process. Because of the need to interview key agency personnel and begin collecting key documentation, pre-studies will probably involve visits to the site of operations.

Another approach towards identifying the key issues is to request personnel involved in the operation to carry out some form of self-evaluation process. This approach has the merit of making the personnel being evaluated feel included and valued in the process. However, such an approach should not prevent the evaluation team from identifying other issues to focus upon, as self-evaluations can miss or avoid underlying problems.

Such an iterative approach requires flexibility on the part of Evaluation Managers. The Terms of Reference may need to be modified or renegotiated during the process. Good quality and frequent communications will need to be established between the Evaluation Manager and the evaluation team. It is likely that a humanitarian assistance evaluation will require more management time than is normally allocated to an evaluation of development assistance.

3.3 Preparing the TOR

Whilst it is tempting to prepare 'minimalist' TOR which simply reproduce the standard evaluative criteria of effectiveness, efficiency and impact, it is important for the TOR to be grounded in the broader questions concerning the nature of the problem and the effectiveness with which it was addressed. Sharpening such broad questions is likely to involve some very specific questions. However, these specific questions must not be set in stone. As the study proceeds so the specific questions may need to be modified or complemented by others. The TOR should therefore be treated as negotiable by the Evaluation Managers and the Evaluation Team. Amending the TOR may require additional work and travel and thus the commissioning agencies must be flexible in their funding arrangements. Evaluation Managers who are unable to provide such flexibility may attempt to deal with the issue through the use of the staged approach or a pre-study 'scoping' phase indicated in Section 3.2. The TOR for the main phase of the evaluation may be finalised in the light of the findings of the pre-study.

The issue of who should be involved in drawing up the TOR is particularly difficult. Broad ownership of the evaluation can be generated if the development of the TOR is an inclusive process. However, different actors frequently have different agendas and these agendas will inevitably influence the TOR. How far these different agendas should be accommodated is by no means fixed. In some situations it may simply not be possible to reconcile the differing agendas. For instance, in the case of neighbouring states which have recently endured hostile relations it may not be feasible to reconcile their different perspectives.

The standard OECD/DAC evaluation criteria of ***efficiency***, ***effectiveness***, ***impact***, ***sustainability*** and ***relevance*** are broadly appropriate for humanitarian assistance programmes. However, their applicability to humanitarian assistance being provided in the context of complex emergencies may be increased through elaboration of certain criteria and the addition of complementary sub-criteria.

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

Effectiveness measures the extent to which the activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criteria of effectiveness is **timeliness** (for if the delivery of food assistance is significantly delayed the nutritional status of the target population will decline). There is value in using it more explicitly as one of the standard criteria because of its importance in the assessment of emergency programmes. Similarly, issues of resourcing and preparedness should be addressed

Impact looks at the wider effects of the project - social, economic, technical, environmental - on individuals, gender and age-groups, communities, and institutions. Impacts can be immediate and long-range, intended and unintended, positive and negative, macro (sector) and micro (household). Impact studies address the question: what real difference has the activity made to the beneficiaries? How many have been affected?

Relevance is concerned with assessing whether the project is in line with local needs and priorities (as well as with donor policy). A recent evaluation of humanitarian assistance replaced the criteria of relevance with the criteria of ***appropriateness*** - the need "to tailor humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly" (Minear, 1994). However, the two criteria complement rather than substitute each other. 'Relevance' refers to the overall goal and purpose of a programme, whereas 'appropriateness' is more focused on the activities and inputs. The expansion of the criteria draws attention to the fact that even where the overall programme goal is relevant - for example, to improve nutritional status - there are still questions to be asked about the programme purpose. Distributing large quantities of food aid may not be the best way of improving nutritional status. Alternatives could include food for work, cash for work, or measures to improve the functioning of local markets. Furthermore, even if distribution of food aid is deemed appropriate, it is still necessary to examine the appropriateness of the food that is distributed.

Sustainability - of particular importance for development aid - is concerned with measuring whether an activity or an impact is likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable. However, many humanitarian interventions, in contrast to development projects, are not designed to be sustainable. They still need assessing, however, in regard to whether, in responding to acute and immediate needs, they take the longer-term into account. Larry Minear has referred to this as ***Connectedness***, the need "to assure that activities of a short-term emergency nature are carried out in a context which takes longer-term and interconnected problems into account" (Minear, 1994). For example, otherwise efficient food distribution programmes can damage roads used by local traders, while the presence of large refugee camps can result in severe environmental impacts in neighbouring areas. Local institutions can also suffer - the high salaries paid by international NGOs can attract skilled staff away from government clinics and schools, leaving the local population with reduced levels of service. Large-scale relief programmes can also have a significant impact on local power structures, for better or for worse.

Coverage - the need "to reach major population groups facing life-threatening suffering wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agendas". Minear alerts evaluators that complex emergencies and associated humanitarian programmes can have significantly differing impacts on different population sub-groups, whether these are defined in terms of ethnicity, gender, socio-economic status, occupation, location (urban/rural or inside/outside of a country affected by conflict) or family circumstance (e.g. single mother, orphan). Programmes need to be assessed both in terms of which groups are included in a programme, and the differential impact on those included. For example, studies have shown that, in Ethiopia in the 1980s, more than 90% of international relief went to government-controlled areas, penalising those in areas of Tigray and Eritrea controlled by insurgent movements (Minear, 1994). Other studies have revealed that single mothers may be disadvantaged when it comes to access to resources, as they are unable to leave children to queue for relief goods. In the case of the Great Lakes emergency, it was found that the coverage of the response varied enormously: refugees and IDPs, and residents in neighbouring IDP camps, were often treated in quite different ways, despite having very similar needs (Borton et al. 1996).

Coherence - refers to policy coherence, and 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. A notable lack of coherence was evident in the international community's response to the Great Lakes emergency in 1994. During the crisis military contingents were withdrawn from Rwanda during the genocide, when there is evidence to suggest that a rapid deployment of troops could have prevented many of the killings and the subsequent refugee influx into Zaire. This was then followed by a huge relief operation. In other instances, donor-imposed trade conditions have been blamed for precipitating economic crisis and conflict, undermining longer-term development policies. Coherence can also be analysed solely within the humanitarian sphere - to see whether all the actors are working towards the same basic goals. For example, there have been instances of one major UN agency promoting the return of refugees to their host country while another is diametrically opposed to such policies.

Finally, there is the important issue of ***co-ordination***. This could be considered under the criteria of effectiveness, for a poorly co-ordinated response is unlikely to maximise effectiveness or impact. However, given the multiplicity of actors involved in an emergency response, it is important that co-ordination is explicitly considered - the intervention of a single agency cannot be evaluated in isolation from what others are doing, particularly as what may seem appropriate from the point of view of a single actor, may not be appropriate from the point of view of the system as a whole.

Given the context of conflict and insecurity, **protection** issues are also critical to the effectiveness of humanitarian action. Where levels of protection are poor it is feasible that the target population of an otherwise effective project distributing relief assistance are being killed by armed elements operating within the project area or even within the displaced persons/refugee camp. Assessment of the levels of security and protection in the area of the project or programme and, where relevant, the steps taken to improve them should be part of all humanitarian assistance evaluations. In those humanitarian assistance evaluations undertaken to date, such issues have often been left out of the study or not adequately covered. Often the lack of familiarity of Evaluation Managers with the issues of security and protection has contributed to such omissions.

International agreements on standards and performance, such as the Red Cross/ NGO Code of Conduct and the Sphere Project, as well as to relevant aspects on international humanitarian law provide international norms against which the performance of agencies and the system may be assessed. The TOR ought therefore to refer the evaluators to such norms.

3.4 Selecting a Team

The complexity of the subject matter and the specialist knowledge required to evaluate different types of intervention and aspects of the programme implies the involvement of a wide range of skills and specialist knowledge. For evaluations of limited scope focusing on just one type of intervention, a small team may be adequate. However, for many evaluations, especially the partial system and total system studies it is likely that additional skills will need to be represented: health; water and sanitation; food and nutrition; shelter; security and protection issues; gender issues, legal matters; media issues; human resource mangement, and so on. Optimum team composition will vary, depending on the particular characteristics of each emergency and the range and type of programs to be evaluated.

Essentially two options are open to Evaluation Managers. One is to assemble a large team and accept the management difficulties that are likely to flow from this. Another is to form a small core team of able generalists, and bring in sectoral expertise where necessary. This might be achieved by establishing an Advisory Group for the core team. The strength of multi-disciplinary teams lies in the differing perspectives that can be brought to bear on the issues. Therefore, regardless of the exact structure of the evaluation team, all team members should get together at several stages of the evaluation process to discuss overlapping issues and conclusions. Provision for such meetings must be included in initial budget estimates.

Experience suggests that a tender process alone seldom produces the optimum team. Good teams are more likely to be produced by a process of short-listing key individuals such as the team leaders, discussing their strengths and weaknesses, and then asking individuals if they are interested, or confining a bidding process to those on the short-list. In other words, a search process may be combined with a tender process.

The position of team leader is a critical one especially where larger teams are involved and team and general management skills are required. Team leaders ought to have strong evaluation backgrounds and experience of evaluating, or operating in, emergency programmes. Team leaders also need to be confident communicators in order to manage the team, to participate in negotiations over terms of reference, to interview senior level key informants and to present the findings once the study has been completed. The writing abilities of the team leader are also important. For instance, an evaluation of operations in Somalia by a bilateral donor ran into difficulties because the technical experts could not write up their sectoral reports and integrate them, so additional resources were needed in order to complete the task. Instead of using the team leader as editor, an alternative option may be to involve a professional content editor early in the process.

3.5 Methods of Working

Preparing the narrative history and ‘baseline’ has to be the starting point for any study. Expansion and modification of the narrative history and ‘baseline’ will probably continue throughout the study as more information is obtained. Whilst documentation will be an important source of information, interviews with the range of actors and members of the affected population will be a vital source too. Arguably, interviews form a more important source than is normally the case with evaluations of development assistance due to the problems of poor record keeping and documentation noted in Section 2.3. Effective management of the results of these interviews is an important determinant of the effectiveness of the team. Generally,

different members of the team should take responsibility for interviewing different individuals in different locations. However, whilst rational in terms of time and travel, such divisions of labour require the development of standard protocols to be used by all members of the team in their separate interviews. Such protocols should include relatively open-ended questions such as "What key lessons did you learn from your experience?" "What in your view were the main strengths of the operation?" and "What would you change if you had to do it all again?" Team members would need to be disciplined in their adherence to these protocols and to writing-up and sharing interview records among team members. The use of laptop computers, e-mail and free-form databases is potentially a very effective means of sharing such information and enabling all team members to contribute to and benefit from the process of constructing the narrative and 'baseline'.

As noted above the strength of multidisciplinary teams lies in the differing perspectives that can be brought to bear on the issues and it is vital that the team has sufficient opportunities to get together at different stages of the evaluation process to discuss overlapping issues and conclusions.

Interviews with a sample of the affected population should be a mandatory part of any humanitarian assistance evaluation. Humanitarian assistance is essentially a 'top down' process. Of necessity it often involves making assumptions about assistance needs and the provision of standardised packages of assistance. Even where time and the situation permits, humanitarian agencies are often poor at consulting or involving members of the affected population and beneficiaries or their assistance. Consequently, there can often be considerable discrepancy between the agency's perception of its performance and the perceptions of the affected population and beneficiaries. Experience shows that interviews with beneficiaries can be one of the richest sources of information in evaluations of humanitarian assistance. The use of Rapid Rural Appraisal and Participatory Rural Appraisal techniques can be very helpful in selecting members of the affected population to be interviewed and in the structuring of the interview. A combination of interviews with individual households, women's groups and open group discussions involving men as well have proven to be very productive in some contexts. However, in the context of recent or ongoing conflicts such a process may need to be modified. Ensuring the confidentiality of some interviews with individuals may be necessary. The deliberate seeking out of those who did not benefit from the assistance available can also be fruitful as it may reveal problems with the targeting and beneficiary selection processes used by the agencies. Ideally, anthropologists familiar with the culture and the indigenous language will undertake this work.

3.6 Writing the Report and Preparing the Recommendations

The writing of the final report, its structuring and language, is pivotal to the whole credibility, competence and independence of the evaluation. It is important that Evaluation Managers appreciate the degree to which the process of writing and analysis are entwined. Adequate time must be allowed for the writing process. A useful rule of thumb is to allow 50% of the time spent in the field by each member of the team.

A draft report should always be prepared and circulated for comment to all those organisations and individuals involved in the evaluation. It is vital for the credibility and competence of the process that a draft is shared widely, that adequate time is allowed for it to be considered and for comments to be received. The complexity of the subject, the inadequacy of documentation, the fallibility of evaluators and the issue of whether an objective truth exists in the context of civil wars, makes this a particularly important requirement in relation to humanitarian assistance evaluations. Experience shows that four weeks is an appropriate time to allow for comments on a draft report. Whilst some comments may be easily dealt with others may require additional investigation. Indeed the sharing of the draft may actually 'smoke out' additional information that had not been shared earlier in the process due to the agencies perceptions of the sensitivity of the information.

An important principle is that whilst evaluators should try to resolve differences of judgement they should not be required to change their judgements on the basis of the comments received. Where differences of opinion between the evaluators and the stakeholder commenters cannot be reconciled, the commenters opinion may be included in a footnote or annex. Given the strength of views often held by those involved in complex emergencies and humanitarian assistance operations, the preservation of the independence of the evaluation team is an important principle to be maintained.

It has already been noted that one of the strengths of a multi-disciplinary team is the differing perspectives it can bring to bear on issues. All team members should, therefore, be involved in discussing the findings and linking these to conclusions. Tensions which may arise between the team leader and individual subject specialists on the nature of the conclusions can be managed more effectively if the whole team is brought together to discuss findings and conclusions. Ideally the team should hold workshops to discuss their preliminary findings on return from fieldwork (where they have not been working together in the field) and then subsequently to discuss comments received on the draft report and any new information provided by the agencies.

Whatever its scope or nature, an evaluation report will maximise its potential impact if it presents its findings, conclusions, recommendations and follow-up sections separately. If readers disagree with the recommendations (or find themselves unable to implement them because of political constraints), they may be able to agree with the findings or conclusions. Comprehensive discussions on the draft report within the target audience of the report is likely to increase their 'ownership' of the report and the likelihood of its acceptance and follow-up.

In any large-scale evaluation, conflict over the nature of the recommendations is probably inevitable. In preparing recommendations, in order to minimise such conflict, there needs to be a clear link between the recommendations and the evidence in the body of the report to support them.

The form in which recommendations should be made is the subject of continuing debate among evaluation specialists. Some would argue that evaluation reports should contain specific, implementable recommendations detailing the actions agencies should take in order to improve future performance. Such recommendations might also spell out who is responsible for implementing each recommendation and who is responsible for monitoring whether this action takes place. This approach has the benefit of making the responsibility for implementation and follow-up clear and reduces the opportunity for organisational evasion and 'fudging'. However, others would urge caution, favouring findings and conclusions over specific recommendations, so as not to burden policy-makers with recommendations that could lead to unimplementable policies. If recommendations are required, an evaluation team might provide policy-makers with options rather than a single recommendation, together with an analysis of expected consequences. Different issues may require different responses. Technical issues may lend themselves to specific recommendations in the final report. In dealing with broader issues it may be useful to deliver the analysis to a workshop of decision-makers and evaluators which negotiates follow-up action.

3.7 Follow-up

Evaluation reports also need to be "sold". Bureaucrats, field officers and agency staff need to be enthused, excited and convinced that the evaluation report is important and should be read. While selling the report is more the responsibility of the management group than the evaluation team, marketing strategies could be included in negotiated follow-up actions in order to help steering committee members sell the evaluation report within their own organisation.

Large, system-wide evaluations raise issues relating to a diverse range of organisations and compliance cannot be compelled. Monitoring of follow-up action is therefore important because it provides for a level of accountability which is otherwise missing. A well-resourced and well-structured monitoring process can strongly influence agencies to account for their response (or lack of it) to the evaluation report. The Joint Evaluation Follow-up Monitoring and Facilitation Network (JEFF), established after the Rwanda evaluation provides a potential model for institutionalising post-evaluation monitoring (JEFF 1997, see Box 10).

It is highly desirable for Evaluation Managers to establish a mechanism whereby decisions relating to the conclusions and recommendations of an evaluation are formally recorded and an explanation provided for where action is to be taken and who is to be responsible. Where no action is deemed appropriate, this would need to be justified. Evaluation departments would then have a basis for monitoring whether the agreed actions are undertaken.

Box 10. Examples of Follow-Up To Evaluations

JEFF consisted of a small group, representing the Steering Committee, the Management Group and the evaluation teams that had been involved in undertaking the Joint Evaluation of Emergency Assistance to Rwanda. It met regularly in the year following publication of the evaluation to report back to the Steering Committee on action taken on recommendations by the bodies concerned. This maintained interest in the evaluation report, and kept up pressure on concerned bodies to demonstrate that they were taking the report seriously.

In 1998 WFP formed a Recommendation Tracking Unit, composed of individuals drawn from senior management who meet to check to see what has happened to the recommendations made in evaluation reports. The Unit's aim is not to ensure that recommendations have been adopted, but rather that recommendations have been debated at the appropriate level. Recommendations are tracked for 18 months following an evaluation, the evaluation process being seen as something that takes place over 2.5 years. Tracking begins with a discussion of what the key issues are, followed by preliminary contacts with the field, and sometimes a pre-evaluation trip to confirm priorities, track down key informants, start looking for secondary data sources and prepare for primary data collection. It ends with a report on whether recommendations have been considered.

Following the 1997 Review of Operation Lifelines Sudan (Karim et al. 1997), the UN agencies that made up the OLS consortium produced a "Plan of Action for Implementation of Recommendations". An initial response to the Review team's recommendations were made, and an action plan drawn up. Every quarter, the consortium monitored progress against the plan of action.

Regardless of whether recommendations are negotiated or independent, specific or general, evaluation reports must be timely if they are to do more than gather dust on the shelves. In spite of the complexity of the issues involved, the speed at which consciousness of the need for emergency assistance fades means that evaluations must take place promptly. Organisations are easily able to ignore evaluation reports which are published two or three years after the events to which they refer.

[Adapted from Hallam 1998]

References

Apthorpe, R. (1996), 'Protracted Emergency Humanitarian Relief Food Aid: Towards Productive Relief' Programme Policy Evaluation of the 1990-95 period of WFP-assisted refugees and displaced persons operations in Liberia, Sierra Leone and Côte d'Ivoire, Rome.

Apthorpe, R. and Gaspar, D. (1996), 'Arguing Development Policy: Frames and Discourses' Frank Cass, London.

Apthorpe, R. and Nevile, A. (1998), 'Managing Emergency Humanitarian Aid Evaluation: Lessons from Experience'. A report on the main findings of a workshop held with AusAID support at the National Centre for Development Studies, Australian National University 13-16 March 1998.

Borton, J. et al. (1996), 'Humanitarian Aid and Effects' Study III of "The International Response to Conflict and Genocide: Lessons from the Rwanda Experience", Danish Ministry of Foreign Affairs, Copenhagen.

Borton, J. and Macrae, J. (1998), 'Evaluation Synthesis of Emergency Aid' EV 613 Department for International Development, London.

Clay, Edward et al. (1995), 'Evaluation of ODA's Response to the 1991-92 Southern African Drought' EV 568, ODA Evaluation Department, London.

Hallam, Alistair (1998), 'Evaluating Humanitarian Assistance Programmes in Complex Emergencies', RRN Good Practice Review No7 September, ODI, London.

JEFF (1997), 'The Joint Evaluation of Emergency Assistance to Rwanda: A Review of Follow-up and Impact Fifteen Months After Publication' Sida, Stockholm and ODI, London

Karim, A, et al. (1996), 'Operation Lifeline Sudan: A Review' DHA-Geneva/University of Birmingham.

Kirkby, J., Kliest, T. & O'Keefe, P. (1997), 'The Complex Emergency in Somalia: UNHCR's Cross Border Cross Mandate Operation. Netherlands Ministry of Foreign Affairs, The Hague.

Minear, 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. September 22.

Ricoeur, P. (1979), 'The Model of the Text: Meaningful Action Considered as a Text', in P. Rainbow and W. Sullivan (eds.), "Interpretative Social Science: A Reader", Berkely, CA: University of California Press.

Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda, (1996), The International Response to Conflict and Genocide: Lessons from the Rwanda Experience, Danish Ministry of Foreign Affairs, Copenhagen.

Further Reading and Resources

ALNAP – the Active Learning Network on Accountability and Performance in Humanitarian Assistance. ALNAP's objective is to improve the quality and accountability of humanitarian assistance programmes by providing a forum for the identification and dissemination of best practice and the building of consensus on common approaches. The ALNAP Secretariat located in the Humanitarian Policy Group, ODI, London. Services include a searchable database of 130 humanitarian assistance evaluations includes a database. www.oneworld.org.odi/alnap/

Anderson, M. B. (1996), Do No Harm: Supporting Local Capacities for Peace Through Aid. Cambridge, USA: The Collaborative for Development Action, Inc

Apthorpe, R. and Atkinson, P. (1999), 'Towards Shared Social Learning for Humanitarian Programmes' ALNAP, ODI, London.

Broughton, B. and Hampshire, J. (1997), 'Bridging the Gap: A Guide to monitoring and evaluating development projects' Australian Council for Overseas Aid, Deaking, Canberra.

Brown University – The Humanitarianism and War Project within the Thomas Watson Institute for International Relations has produced a range of studies on humanitarian assistance issues including reviews of programmes and studies on particular issues such as coordination and sanctions. www.brown.edu/Departments/Watson_Institute/

Chambers, R. (1992), Rural Appraisal: Rapid, Relaxed and Participatory, Institute of Development Studies Discussion Paper No. 311. Brighton: IDS

Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief, (1994), Relief and Rehabilitation Network Paper 4. London: Overseas Development Institute.

Court of Auditors Special Report No 2/97, concerning humanitarian aid from the European Union between 1992 and 1995 together with the Commission's replies, Luxembourg (for further details see http://europa.eu.int).

de Veer, T. (1996), Assessment of Experiences, Attitudes, Beliefs, Developments and Ideas Among Implementing, Donor and Support Agencies with Respect to Monitoring and Evaluation of Water and Sanitation Programmes in Camps for Refugees Or Displaced Persons. Leiden: Emergency and Humanitarian Aid Bureau of the Netherlands Ministry of Foreign Affairs.

Dabelstein, N. (1996), "Evaluating the International Humanitarian System" in Disasters Vol. 20, No. 4, ODI, London.

Darcy, James (1997), Human rights and international legal standards: what do relief workers need to know?, Relief and Rehabilitation Network Paper 19. London: Overseas Development Institute.

Donini, A., Dudley, E. & Ockwell, R. (1996), Afghanistan: Co-ordination in a Fragmented State: Lessons Learned Report prepared for the United Nations Department of Humanitarian Affairs. Geneva: UNDHA.

ECHO Operational Manual for the Evaluation of Humanitarian Aid, available from ECHO, Brussels.

Gasper, D. (March 1998), Notes on the Logical Framework Approach and the Planning and Evaluation of (Complex) Emergency Relief, The Hague: Institute of Social Studies

Gosling, L. and Edward, M. (1995), Toolkits: A Practical Guide to Assessment, Monitoring, Review and Evaluation, Save the Children Fund, London.

Hallam, A. (1996), Cost Effective Analysis: A Useful Tool for the Assessment and Evaluation of Relief Operations. Relief and Rehabilitation Network Paper 15, April 1996. London: ODI.

Harvey, P. (1997), PRA and Participatory Approaches in Emergency Situations: A Review of the Literature and ACTIONAID's Experience. London: ACTIONAID.

Lautze, S., Jones, B. and Duffield, M. (1998), 'Strategic Humanitarian Coordination in the Great Lakes Region 1996-1997'. An Independent Study for the Inter-Agency Standing Committee, Office of the Coordinator or Humanitarian Assistance, UN, New York.

Lindahl, C. (March 1998), The Management of Disaster Relief Evaluations: Lessons from a Sida evaluation of the complex emergency in Cambodia, Solna, Sweden (email: mpi@swipnet.se)

Mansfield, D. (1996), Evaluation: Tried and Tested? A Review of Save the Children Evaluation Reports. London: Save the Children Fund

Martin, L. (Jan 1998), Planning and Organising Useful Evaluations, UNHCR, January.

Minear, L. (1998), 'Learning to Learn'. Discussion paper presented to the Seminar on Lessons Learned on Humanitarian Coordination, 3-4 April, Stockholm.

MSF Holland Monitoring and Evaluation Unit (1997), Formulating Results for Planning, Monitoring and Evaluation: A Practical Approach. Amsterdam: MSF.

MSF Holland (1996), Mini Symposium on Evaluation and Impact Studies of Humanitarian Relief Interventions. Final Report of the Symposium held in Amsterdam, Netherlands, November 28 and 29, 1996. Amsterdam: MSF.

Neefjes, K. (1999), 'Participatory Review in Chronic Instability: The Experience of the Ikafe Regugee Settlement Programme, Uganda' RRN Network Paper 29, ODI, London.

OECD/DAC (1997), DAC Guidelines on Conflict, Peace and Development Cooperation. Paris: OECD.

Reliefweb - a large website maintained by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) with the intention of strengthening the response capacity of the humanitarian relief community through the timely dissemination of reliable information on prevention, preparedness and disaster response http://wwwnotes.reliefweb.int/

RRN – Relief and Rehabilitation Network. A mechanism for the exchange and dissemination of information relevant to the professional development of those engaged in the provision of humanitarian assistance comprising over 1,100 members in 70 countries. Produces regular Newsletters, Network Papers and Good Practice Reviews. www.oneworld.org/odi/rrn/

Slim, H. & Mitchell, J. (1997), The Application of RAP and RRA Techniques in Emergency Relief Programmes in Scrimshaw, N. S. & Gleason, G. R. (Eds) (1997), Rapid Assessment Procedures: Qualitative Methodologies for Planning and Evaluation of Health Related Programmes. Boston: INFDC.

Sørbø, G. M., Macrae, J. & Wohlgemuth, L. (1997), NGOs in Conflict: An Evaluation of International Alert. Bergen: CHR Michelsen Institute.

White, L. (1994), Policy Analysis as Discourse, Journal of Policy Analysis and Management, Vol. 13, No. 3, pp. 506-25.

