

**Partners in Development Evaluation –Learning and
Accountability
Joint French and DAC/WP-EV Workshop
Paris, 25 –26 March 2003**

Practice makes perfect? ALNAP’s assessment of the quality of 127 evaluations of humanitarian action, 2000-2.

Presentation by John Mitchell, ALNAP Coordinator, London, UK¹

Contact: j.mitchell@odi.org.uk; www.alnap.org

Summary: This paper reports on the findings of a meta-evaluation of 127 evaluations of humanitarian action carried out under the auspices of ALNAP’s *Annual Reviews* for 2001, 2002 and 2003. The evaluation reports come from across the humanitarian sector and can be considered representative of the work of the sector as whole.

Strengths of evaluation of humanitarian action include mainstreaming of the DAC evaluative criteria, and sustained attention to coordination, and management and human resource issues. There are however a number of troubling weaknesses, in particular: poor attention to evaluation of rights based issues including adherence to international standards, protection and gender equality; inadequate methodologies including weak consultation with primary stakeholders; lack of transparency in description of methods used; and failure to promote adequately evaluation use. Overall these weaknesses tend to undermine the credibility of many evaluations.

1. Introduction to ALNAP and the ALNAP *Annual Review*

The origins of ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action) lay in the recognition that the evaluation of humanitarian action (EHA) needed greater capacity, coordination, opportunity for networking and sharing of information. However, recognition is one thing and action another, and the catalyst for ALNAP’s creation was the joint evaluation of emergency assistance to Rwanda (ODI 1996), the first, and to date only, system-wide evaluation of humanitarian action, which brought together a set of evaluation actors from a number of key agencies, and eventually led in 1997 to ALNAP’s formation, with a Secretariat located at the Overseas Development Institute in London.

ALNAP is an international interagency forum working to improve learning, accountability and quality across the Humanitarian Section. It has 51 Full Member Organisations, including all relevant UN agencies, the International Red Cross/Red Crescent Movement, bilaterals, NGOs, academics, research centers and consultants. It also has 287 Observers. ALNAP is governed by a Steering committee of eight Full Member Representative selected to maintain balance of representation between organizational types. As far as non-governmental evaluation coordination and networking mechanisms go, ALNAP is probably unique.

¹ The contribution to this paper of Tony Beck, consultant to ALNAP and one of the authors of the ALNAP *Annual Review*, should be noted.

ALNAP's activities are funded by Full Members on a voluntary basis. Main current areas of work are:

- Network activities, in particular Biannual meetings for sharing information and building consensus on ways to improve EHA.
- The *Annual Review Series*, to which the paper will return in the next section.
- The Evaluative Reports Database, a fully searchable internet database which holds about 400 evaluative reports.
- Training modules and guidance material.
- Interest group activities, currently a global study on beneficiary consultation and participation, and a Learning Support Office.

2. Introduction to the ALNAP meta-evaluation exercise

Since 2001 ALNAP has published an *Annual Review* of EHA which has included a thematic review (Kosovo in 2001, learning in 2002, and monitoring in 2003), and a synthesis and meta-evaluation of reports provided to the Evaluative Reports Database for that year. The total number of reports reviewed over the three year period is 165, of which 127 were included in the meta-evaluation.

The purpose of a meta-evaluation is to assess the quality of a thematic group of evaluations and subsequently to suggest improvements. As Lipsey points out (2000: 212):

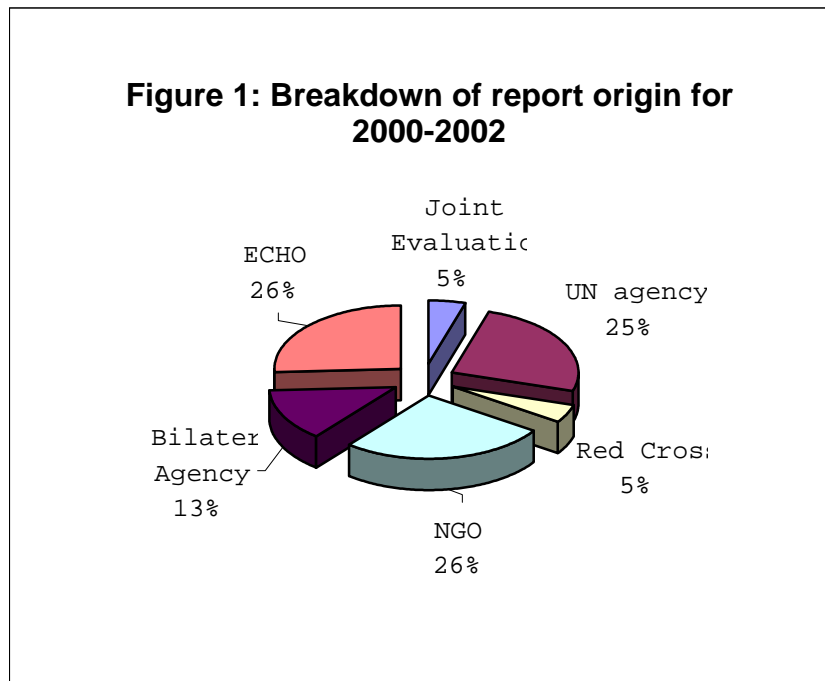
Meta-analysis and other forms of systematic synthesis of evaluation studies provide the information resources for a continuous improvement program for evaluation practice itself. By examining the patterns and relationships revealed by meta-analysis, an evaluator will better understand what program characteristics, outcome domains, and research methods are most likely to be important for a particular evaluation effort. As new evaluation studies are completed and added to cumulative syntheses, the knowledge resources of the evaluation field will become richer and more differentiated and their potential contribution to practice, in turn, will become more useful.

As ALNAP has been carrying out meta-analysis of EHA for three years it is in a position to draw broad conclusions about its quality. The meta-evaluation is carried out by two assessors using the ALNAP Quality Proforma (downloadable from the ALNAP website). This paper gives only a skeleton background to the Proforma so as to make the results to which the paper turns in a moment understandable – further details can be found in the *Annual Review*. The Proforma is used to assess the quality of seven areas in the evaluation report: terms of reference; approach and methods; contextual analysis; analysis of the intervention from planning to monitoring, including cross-cutting themes such as gender equality, protection and international standards; assessment of evaluation practice, including conclusions and recommendations; and report legibility and accessibility. Each area is rated on a four point scale: poor, unsatisfactory, satisfactory and good.

The analysis below presents strengths and weaknesses in EHA over the period 2000-2002 in key EHA areas. For convenience the breakdown is given by satisfactory/unsatisfactory ratings only. Good practice is also highlighted.

2.1 The sample

The breakdown of the sample by agency over three years is shown in Figure 1. Between them the UN, NGOs and ECHO provided 77 per cent of the evaluations received by the Evaluative Research Database.



Bilateral donors and ICRC could be better represented and both need to commission more evaluations and/or make these publicly available.²

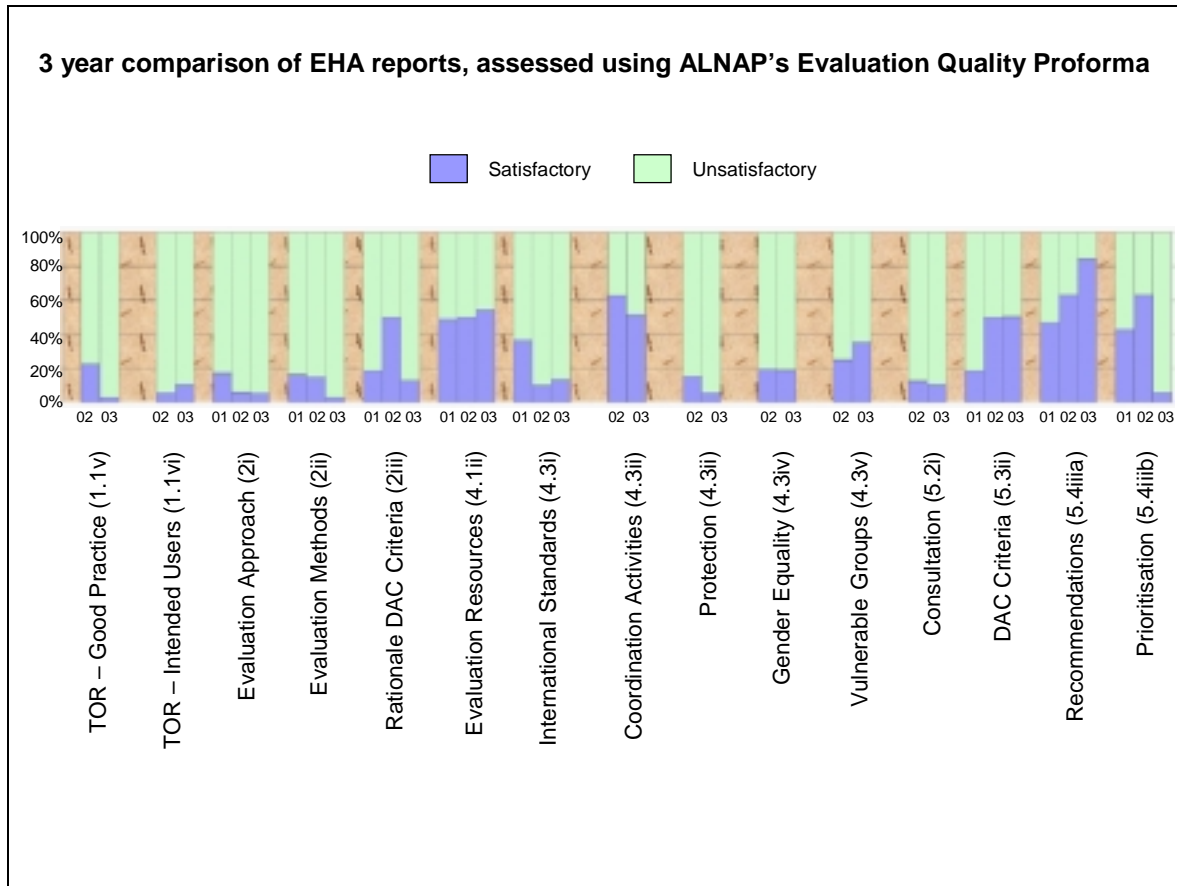
3. Findings from the three year comparison

3.1 General findings

Figure 2 provides an overview of evaluation performance in 15 areas for all evaluations analysed. Findings are grouped together in this way so as to allow an overview of performance in all the areas considered. After the Figure, individual areas are discussed, with reference made to the Figure as appropriate.

² ICRC reports may be subject to greater confidentiality issues than other EHA.

Figure 2³



3.2 Evaluation terms of reference

Two areas are covered here:

- The quality of the terms of reference (TOR) statement on expectation of good practice in approach and method, eg, application of DAC criteria; reference to international standards including international law; multi-method approach – ie, quantitative and qualitative; consultation with key stakeholders to inform findings, conclusions and recommendations; and gender analysis (column 1 in Figure 2)
- The quality of TOR statement on intended use and users of evaluation outputs (column 2 in Figure 2).

Evaluation reports were found to be weak in both of these areas. In general reports did not specify adequately the key methodological areas that evaluators should use. It was also rare for TOR to outline clearly the intended use of evaluation reports; failure to do this adds to the

³ Because of lack of space, not all areas in Figure 2 are included in the analysis below. Comparative data was not available for all three years. Some decline in performance between 2002 and 2003 can be explained by changes in definition of what was 'satisfactory' in the Proforma. Figures in brackets in Figure 2 refer to the section heading of the Proforma for the *Annual Review 2003*.

likelihood that the findings of these reports will not be fully used. Some good practice was evident, for example an evaluation of WHO's intervention in East Timor which noted (WHO 2001: p37):

WHO will make efforts to make partners and interested parties aware of the report and make the report accessible to them. Comments on the report will be requested from collaborating partners. The report will be made available on the WHO web site.

The report will be discussed at WHO Dili office, SEARO and HQ. The conclusions and recommendations will be formally commented upon linked to the places and responsibilities to which they apply. Decisions will be made regarding follow up, compliance will be reported 6 months later.

Another good practice example is the evaluation of WFP's intervention in the Great Lakes (WFP 2002), which notes that: the report will be presented to WFP's Executive Board; key recommendations arising from the evaluation will be used in the preparation of a Management Response Matrix which will outline how the WFP Regional Bureau in Kampala intends to follow up on the evaluation's key findings and recommendations; and dissemination through WFP's website and a publicly available summary.

3.3 Evaluation approach and methods

Two areas are covered here:

Appropriateness of the overall evaluation approach (column 3 in figure 2). 'Approach' here means the wider conceptual framework used and evaluation tradition being drawn upon, such as accountability-oriented, utilization-focused, or empowerment evaluation approaches.

Rating in this area was generally unsatisfactory, in particular in 2002 and 2003, where 80 per cent and over of reports were rated as unsatisfactory. EHA does not generally draw on wider evaluation theory. The implications of lack of attention to wider evaluation discourse are discussed in detail in the *Annual Review 2002*, and include: a lack of conceptual direction for the evaluation; an inability to rationalise why a particular evaluation methodology has been selected; a fall-back on 'standard' evaluation techniques with little experimentation; and lack of attention to causality.

b. Appropriateness of the evaluation methods selected, and planned application of the DAC criteria (column four in Figure 2) A significant majority of reports are also rated as unsatisfactory in this area. Most reports note only basic details of the methodology used, which in turn undermines the credibility of findings.

3.4 Quality of the evaluation of agency's management and human resource practices (column six in figure 2)

Areas considered were evaluation of the level of experience/expertise of field staff; recruitment procedures; staff turn over; field/HQ relations; management and development of national staff; briefing and debriefing procedures; training and learning practices; and security. This is a consistent strength in EHA, with over 50 per cent of evaluations rating as satisfactory or better in each year. Evaluators of humanitarian action clearly feel comfortable with assessment of

management and human resource issues, but this in turn may negatively impact their evaluation of other key areas such as protection and adherence to international standards.

3.5 Cross-cutting themes (columns 7 to 11 in Figure 2).

The cross-cutting themes that consistently scored well are coordination and, to a lesser extent, attention to vulnerable and marginalized groups. In the other three areas of gender equality, protection, and adherence to international standards, reports consistently performed poorly. The link between these areas are that they deal with rights based issues; these are the issues that are most often left out of evaluation TOR and with which evaluators appear to feel least comfortable. Protection is particularly poorly covered, with 92 per cent of year 2003 reports and 79 per cent of 2002 reports assessed as unsatisfactory or worse. The Red Cross/Red Crescent Code of Conduct and the SPHERE Standards are also not generally used.

Good practice does however exist in each of these areas. Over the three year period almost all WFP reports were rated as satisfactory or better in evaluation of gender equality. In its thematic review of its Protracted Relief and Rehabilitation Operations programming modality, made up of seven country/regional reports, attention to gender equality is mainstreamed throughout the reports as well as being included in a separate section, the latter often being quite substantial. Also of note is an Annex in each report which contains a checklist on 'Meeting the WFP commitments to women and mainstreaming a gender perspective'. The quality of reporting also suggests that WFP has made a commitment to hiring evaluators who have relevant skills in assessing gender equality. Overall, this is probably the most sustained attention to the evaluation of gender equality in EHA to date.

UNHCR reports are strong on the evaluation of protection, for example the reports on protection of children (UNHCR 2002), the evaluation of support to IDPs in Angola (UNHCR/Danida, 2001a), and of a firewood project in Kenya aiming to reduce rape (UNHCR 2001b). The evaluation of Disasters Emergency Committee (DEC) agencies after the Gujarat earthquake provides a detailed analysis of the response against both the Red Cross/Red Crescent Code of Conduct and the SPHERE Standards (DEC 2002).

3.6 Consultation and participation of primary stakeholders (column 12 in figure 2)

In order to rate as satisfactory, evaluations were required to both undertake adequate consultation and promote participation by primary stakeholders, and outline the nature and scope of this. Weakness in this area, with 73 per cent of reports rating unsatisfactory or worse in 2002 and 89 per cent in 2003, is a serious shortcoming of EHA. However, it is clear that many evaluators are talking to primary stakeholders, and some of the reports are peppered with their quotations or comments, but the method used is not included. Why evaluators do not detail these interviews adequately is unclear. It may be that:

- they do not see the relevance of including this information in the reports, thinking perhaps that it will lead to information overload;
- they are not aware of the importance of explicitly comparing the perspectives of different stakeholders to add credibility to the evaluation findings;
- much of the focus of EHA is on intra-institutional matters and field trips to project sites are rushed and given low priority.

However, consistently consulting with primary stakeholders, cross-referencing this with other perspectives, and detailing the nature and scope of consultation, will go a long way to overcoming one principal problem with EHA, its failure in many cases to establish credibility of evaluation methods.

The DEC provides consistent good practice in this area, as highlighted in Box 1.

Box 1 Good practice in consultation with primary stakeholders

Disasters Emergency Committee (DEC) evaluations have consistently consulted with primary stakeholders. The evaluation of DEC agencies' performance in their response to the 2001 Gujarat earthquake is an excellent example of the levels of consultation that can be achieved:

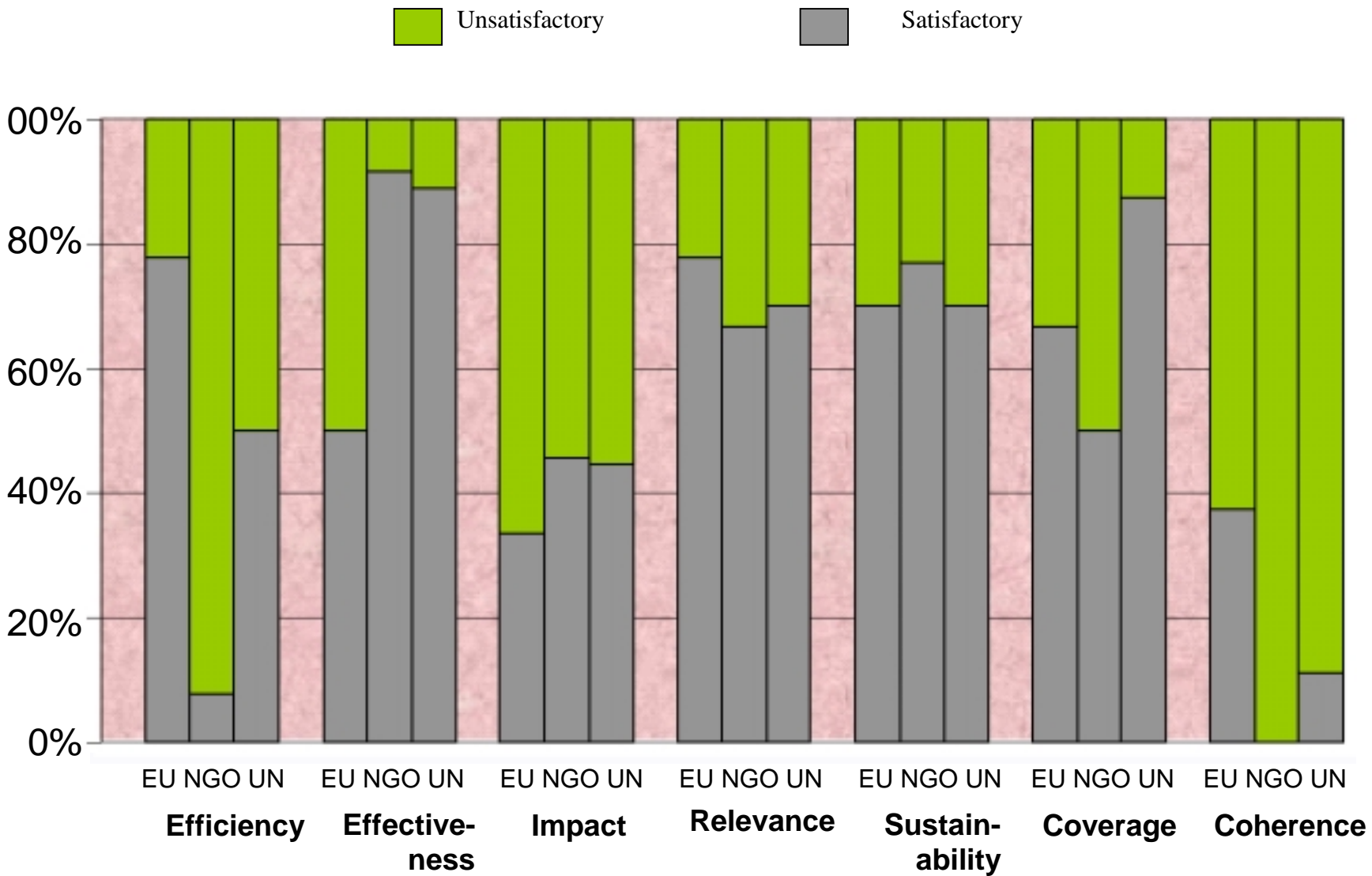
- The evaluation notes the importance of attempting to empower communities through evaluation approaches that seek their active participation.
- The evaluation team included an Ahmedabad based disasters institute, the Disaster Management Facility (DMF). DMF organized and conducted a survey covering 50 villages, and interviews with over 2,300 people. The inclusion of national researchers and consultants is a regular feature of DEC evaluations, unlike most other EHA.
- Interviews and focus groups were carried out using state of the art participatory methodologies, and there was considerable attention paid to the location of consultation exercises in order to encourage the participation of as diverse a cross section of the community as possible.
- Specific attempts were made to include 'missing voices', including low status communities, the poorly educated, widows, women, the disabled and sick, those living on the outskirts of communities, and working in nearby towns during the day. Timing and location of exercises and follow-up interviews attempted to include these groups.
- The methodology is detailed extensively.
- Quotes and comments from primary stakeholders are used effectively throughout the report to substantiate key points.

Source: DEC (2001)

3.7 Application of the DAC criteria (column 13 of Figure 2)

This area of the Proforma assessed whether the DAC criteria had been applied adequately, and this was found to be one of the strengths of reports. Evaluators have developed the capacity to use the DAC criteria, and most EHA is now organized around these criteria.

Figure 3 Application of DAC Criteria, 2003



The analysis of the 2002 evaluation set included for the first time disaggregated assessment of the DAC criteria, presented in Figure 3, by criteria and type of agency. Reports were particularly strong in the evaluation of effectiveness, relevance/appropriateness, and sustainability/connectedness, where there is range of good practice, suggesting that use of these criteria has been mainstreamed into EHA; and to a lesser extent in evaluation of coverage. Efficiency, impact and in particular coherence fared less well. Across the different groups of agencies, NGOs can be seen to be weakest in evaluation of efficiency and coherence. Of note is the relative success in evaluation of effectiveness and to a lesser extent impact; as many agencies move to results based planning there does appear to be greater attention to outcomes and impacts than previously.

4. Cross-agency analysis

Finally, were there any marked differences between agencies as far as evaluation quality is concerned? All actors had their strengths and weaknesses, some of which can be highlighted as follows:

- ECHO and NGOs did relatively well in assessment of coordination, management and human resources and attention to the vulnerable, but relatively poorly in detailing evaluation process and paying attention to adherence to international standards, protection and gender equality.
- Overall, UN agencies performed better in most areas than the other actors in EHA, with aggregate scores for the 3 year period that were highest in almost all areas of the Proforma. This relates in particular to strong evaluation performance by WFP and UNHCR.

5. Conclusions on areas needing to be strengthened

Four points can be highlighted from this analysis:

- **EHA needs stronger methodologies:** Well developed methodologies, including adequate attention to rights based issues, and employing multi-method approaches including consultation with primary stakeholders, are relatively rare in EHA.
- **Need for greater transparency:** Evaluators could do more to outline the evaluation methods used, for example the nature and scope of consultation with different stakeholders, or the reason for selection of sites for field visits. Producing a credible description of methodology would in the majority of cases require only a little more effort – for example noting the numbers of primary stakeholders consulted, broken down by sex and other salient social characteristics, where they were consulted, and the methods used for consultation (questionnaire, focus group etc).
- **Lack of experimentation:** EHA follows a standard ‘model’ with little experimentation and almost no acknowledgement of the weaknesses and strengths of the approach followed. One example of experimentation – the DEC large scale ‘opinion survey’ – is given in Box 1.
- **Lack of attention to social process and causality:** EHA tends to examine what happened rather than why it happened, that is to focus on results rather than their causes.

Linked to this, there is often inadequate attention to social process, for example power relations or gender relations.

The meta-evaluation process has been useful in identifying strengths and weaknesses, and ALNAP's ongoing work in training and development of evaluation guidance aims to support the development of capacity in each of the areas above.

Bibliography:

(DEC 2001) *Independent Evaluation of Expenditure of DEC India Earthquake Appeal Funds: Volume One - Executive Summary; Volume Two - Full Evaluation Report; Volume Three - Sector Reports* London: Disasters Emergency Committee.

Lipsey, M. (2000) 'Meta-analysis and the Learning Curve in Evaluation Practice.' *American Journal of Evaluation*, 21 (2), 207-13.

ODI (1996) *Humanitarian Aid and Effects*. London: Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda, Study 3.

UNHCR (2002) *Meeting the Rights and Protection Needs of Refugee Children: An independent evaluation of the impact of UNHCR's activities*. Geneva: UNHCR.

UNHCR/Danida (2001a) *UNHCR's Programme for Internally Displaced People in Angola: A joint Danida/UNHCR review*. Geneva: UNHCR.

UNHCR (2001b) *Evaluation of the Dabaab firewood project, Kenya*. Geneva: UNHCR.

WFP (2002) *Full Report of the Evaluation of the GREAT LAKES Regional PRRO 6077 and PRRO 6077.1 (WINGS 10062.0): Food Aid for Relief and Recovery in the Great Lakes Region*. Rome: World Food Programme.

WHO (2001) *Review of WHO's Emergency Response in East Timor*. Dili: WHO.