

evaluative reports from even 'ethno-nationalism', a term that appears only once or twice in the 50 or so sources. There is uncritical and unanalytical reference to ethnicity, as if this were all that social and political analysis of the local and national context required.

The paucity of training is an important factor emerging as part of the story told by the evaluations, but is also an issue for evaluators in respect of the skill composition within evaluation teams.

3.4.2 **Humanitarian Agencies, Actors and Acts**

All the Kosovo evaluations agree that evaluations should focus on more than just operations (and agencies). However, they are often discouraged or prevented from doing so by their commissioning organisations.

What an organisation is and what it does is never solely defined by mandate. For one thing, few such mandates are descriptive, rather, they are exhortatory. For another, organisations have staff, resources, histories, problems, strategic objectives and so on, each an influencing factor in what an organisation is and does. This meta-evaluation takes the view that other levels of analysis are equally important, in particular the analysis of individual actors and their acts.

As shown earlier, individual actors (agency personnel) may not be familiar with, let alone observant of, the codes to which their organisation subscribes, or even their organisation's mandates. A focus on acts allows for a comprehensive assessment, whatever the mandate and ethical code of an organisation. Even where mandates and codes are known, how they are understood and interpreted may be more influential to what is said and done than what a mandate says and intends. The evaluative reports show that beliefs vary considerably, even within the same organisation. Saying and signing is not necessarily believing and practising. The importance of circumstance and situation is above all not to be written-off in emergencies – hence the importance of the post-modern approach in this chapter.

Humanitarian organisations consider themselves to be non-commercial but nevertheless perform certain commercial acts such as surviving as organisations, through marketing and public relations. Similarly, commercial organisations (in this instance the military) have non-commercial acts to perform. In other words, not everything a relief organisation imagines, says, believes, and

does, is relief and not everything the military imagine, say, believe, or do, is concerned with combat. There are multiple realities that determine what a particular organisation does in a particular situation, which may throw up concerns about the degree of 'prostitution' of humanitarianism. A debate about just humanitarian versus non-humanitarian organisations (i.e., the military) misses out too much to be practical or intellectually acceptable.

In sum, it is important for evaluations of humanitarian action not to accept stated objectives, policies and principles as givens. A more thorough and nuanced approach will be able to account for the realities of practice in relation to principle.

3.4.3 **Policy Evaluation Contrasted with Project Evaluation**

Evaluation of humanitarian action tends to lean heavily towards project evaluation, this being the primary reference of the increasingly standard criteria: appropriateness, efficiency, impact, coverage, connectedness and coherence (yet to be appraised for their value with regard to the evaluation of humanitarian policy).

Unlike previous evaluations, the Kosovo set of reports are notable for the extent to which they combine a policy and project/operations focus, undoubtedly due to the policy issues thrown up by NATO's involvement.

However, little more than introductory considerations on policy analysis were included in the 1999 OECD *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* (pp 24–25). A key requirement of the evaluation profession will be to determine what the evaluation of project and policy have in common, as well as their differences.

While both policy and project (i.e., programme) evaluation requires recourse to objectives and indicators, these are not always provided by the agency project or programme under evaluation. The WFP evaluation remarks that: '[T]he difficulties in evaluating [WFP's response] can be found in the lack of specific objectives and assumptions formulated, as well as in the lack of proper indicators to measure achievement and assess the assumptions' (WFP, 2000b: p2). A frustration echoed by others.

Induction-led evaluation expects to find the criteria it needs for its own purposes provided, as it were, on a plate. Where this is not the case, if evaluators cannot find ways to devise and provide these for

themselves, they are exposed to the charge of not getting their job done. The responsibility, however, must be shared by the agency being evaluated.

3.4.4 **Lesson Learning and Lessons to be Learned**

Lesson learning is stated as a principal objective of most of the Kosovo evaluations (also most of the policy review documents and workshop reports). They return again and again to the theme that fundamental lessons for strategic planning, the organisation and management of the humanitarian response, and indeed its monitoring and evaluation – which may well have been recommended by previous evaluations – remain to be learned.

Most of the Kosovo evaluations consider that some of the shortcomings of the humanitarian response are due to *force majeure*. The circumstances of this conflict are seen to be exceptional in various regards therefore special allowance should be made. This meta-evaluation has no general quarrel with special pleading, provided what is considered to be specific and perhaps excusable is clearly distinguished from what is not. However, what it does seek to raise is whether, in part, typical evaluation practice is as much to blame for lack of lesson learning as are the evaluated organisations themselves.

Shortcomings in present and past evaluation practice are certainly part of the total picture to be explored. Terms of Reference are typically overloaded. Divisions of labour in evaluation teams tend to be on a sectoral or functional basis only. Conclusions and recommendations of the ‘single best’ form – i.e., not menus of options with the advantages and disadvantages of each option spelt out – is the dominant practice. But on top of this comes the handicapping trend of seeing lesson learning as simply filling knowledge deficits (a practice all too familiar in evaluation of economic development cooperation), uncomplemented by serious attention to knowledge management and implementation.

Current evaluation practice, team composition and so on, is singularly ill-designed and ill-equipped to address the lesson learning purpose that it sets, and has had set for it, unless it can shift its focus to organisational (institutional) learning. This learning is usually a task that agencies set for themselves, but to which evaluation could make a useful contribution. Unless the evaluation process recognises its role in relation to institutional learning, it will lose its status.

3.4.5 **Opaqueness in Evaluation Methodology**

Lack of methodological rigour results in impressionistic findings that are hard to defend, and undermine the reputations of humanitarian evaluations. Giving sufficient attention to methods and methodology is critical to both the quality of, and confidence in, evaluations.

Methodology (M) is sometimes neglected as an assumed 'given', understood by all. In reality there is no common approach or understanding, and denial of any use of methodology in the undertaking of an evaluation is not unknown. It is critical to recognise methodology's multiple layered nature. The following four layers are suggested: M-1, M-2, M-3 and M-4.

M-1 is the layer covered by most of the Kosovo reports. This includes the timetable, who was where, when and why during the mission, the basis for the division of labour in the team and the time allocation in the field and other locations. Although extremely limited in analytical value, it reveals the basics of how an evaluation proceeded and whether it spent enough time in the field and HQ

The M-2 layer should include policy analysis methods, evaluators' involvement in finalising the Terms of Reference, team management issues, established networks, and how difficult relations in the field were addressed – whether with representatives of governmental and other agencies, or of the client organisation.

The M-3 layer should cover conceptual and analytical or methodological frameworks for the study, including precise field methods for data collection, the criteria against which field information was assessed and judgements reached. It should also outline the extent to which evaluators felt the methodology worked and implications for findings, conclusions and recommendations.

Clarity of criteria is particularly important where the influence of team members, as optimists or pessimists, can be both subjective and influential – i.e., is it assessed as a positive that a programme achieved so much or as a negative that it failed to achieve everything. Such assessments should not be totally subjective.

Only one evaluation mentions methodology in this way: 'Our assessment is qualitative. While a method of marking that awards points on a scale, of say 1 to 10, may be appropriate for a simple

project evaluation, it is much too simple for evaluating agency operations in a complex historical context. ... The standard tools used are those of historical and public policy analysis' (UNHCR, 2000a: p3). But information about which tools and what they yielded, or not, compared with others, is not given, nor can it be deduced from the text of this example.

The final, M-4 layer, ought to guide the critical reader as to how the evaluators drew conclusions from findings and recommendations from conclusions. These reports do not. Given that the client (and others) is likely to judge recommendations not so much in terms of 'true or false' but rather 'feasible or costly', a methodology that is transparent would tell us the parameters used to reach such recommendations.

Despite the associated difficulties, it is all the more important in the face of complex field situations, multiple actors, lack of data, variable translation of field interviews, limited access to the field and beneficiaries, and difficulties with attribution, to give prior attention to methods and methodologies. However, the flexibility needed to respond to issues and ideas emerging from the fieldwork, and to cope with resistance met in the field, must be maintained. A two-stage fieldwork approach allowing an initial field visit or tour of the study area can help focus the development of appropriate methods and methodologies.

Short-term consultancy is unlikely ever to compare favourably with research, committees of enquiry, select committees, independent commissions and so on, in terms of its hard data findings. But evaluations are normally expected to yield *some* hard findings. One evaluation claims to believe, but unfortunately fails to demonstrate, that: '[W]ith regard to final results, the most important finding of the evaluation is the high level of satisfaction of beneficiaries' (ICRC, 2000: p 1). This level of aspiration is exceptional in evaluation being more typical of research.

Greater transparency in addition to less overloaded Terms of Reference is needed. It would behove everyone to develop frames and discourses that permit greater focus on conclusions and recommendations, including the provision of a scenario-like menu of options in the recommendation section.

The comparative advantage of evaluation as a lesson learning tool needs to be reconsidered from time to time. It is the functions, purposes and methods of evaluation, and not the genre itself, that

is at issue – including the extent to which evaluation should be linked in practice with other approaches to learning (and participation and ownership).

3 4 6 **Independent Evaluation: Achieved, Necessary, Relevant?**

To what extent is the independence of the evaluation really necessary? And what does it seek to achieve? Independence may well not have been achieved even where evaluations claim that status

Is independence more strategic than other considerations for credibility – such as expertise, authority, or sometimes nationality? In other words, does it function more as a ‘credibility-placebo’ where it is no guarantor, in and of itself, of quality of output?

The reality of independence is that it can be distinctly tenuous, particularly where the like-by-like approach to commissioning extends to evaluators carrying out commissioned evaluations for past employers.

To deconstruct ‘independent’ further – i.e., independent from what? – the configuration of particular disciplines, ideologies, sectors etc. should be considered, as well as personal values and premises. To whom does the evaluator report? An agency’s top management or board, a beneficiary union, an ombudsman, or another sort of watchdog? These issues are not necessarily unrelated to the choice of one consultant over another.

That those stakeholders taken seriously are not generally beneficiaries or their representatives, or the public that contributed money to appeals, is another important part of this picture. The stakeholder might find token reflection in evaluation team composition and the M1 features of methodology, but when even this tokenism is missing, greater concern arises in respect of what is done in the name of independent evaluation

This chapter’s principal author believes:

- mixed teams of insiders and outsiders are more likely to provide credible analysis and recommendations (most Kosovo evaluations fall into the category of outsiders only); and,
- in some circumstances, that invited consultancy has a better chance of being read and heard by the client than an unsolicited paper in a professional journal.

A USAID 'evaluation tip' states: 'If objectivity and credibility are key requirements, an external evaluation may be the appropriate choice, whereas if stakeholders' ownership and acting on it are priorities, more collaborative or participatory approaches are usually better' (USAID CDIE, 1997: p3).

Given that complex emergencies inevitably involve many agencies and that interagency relations are likely to be an important factor in any humanitarian response, more interagency evaluations should be undertaken with their potential for increasing transparency – possibly a more critical characteristic than token independence.

3.4 7 **Paucity of Social, Cultural and Political Analysis**

Applying a humanitarian label should not obscure the political causes or consequences of a complex emergency. According to a recent *Disasters* journal article: 'It is now part of received wisdom that humanitarian assistance in conflict and post-conflict situations may be ineffective or even counterproductive in the absence of an informed understanding of the broader political context in which so-called "complex political emergencies" occur' (Cliffe & Luckham, 2000). The reality of the majority of the Kosovo evaluations is a poverty of social, cultural and political analysis.

This virtual void has huge ramifications in terms of what is learned and not learned about the conflict and conflict-affected populations. Economic analysis of the cost-effectiveness of operations is for the most part also missing, despite the fact that regional economic poverty is widely considered to have had an impact on every aspect of the response – including dilemmas for relief exit strategies.

Is this analytical *un*intelligence simply due to a lack of appropriate skills on the part of evaluators and programme managers? In addition to increased training, the growing rights-based approach to relief might have the invigorating effect of forcing evaluators to consider local people, their perspectives and contexts.

A final area of consideration, supported by some of the evaluations, are factors influencing the effectiveness of faith-based agencies' relief work as compared with non-faith based ones. A number of key dimensions should emerge from such a comparative study, though some might relate more, for example, to the issue of existing and perhaps continuing presence and national staff than to faith and ideology.

3.4.8 **Programme Outputs, Outcomes and Impacts**

As could be expected from evaluations that, for the most part, display no social learning with regard to the affected populations, they offer little by way of impact analysis. Some note this themselves. Others proceed as if anecdotal evidence were enough. On the whole the evaluations do not meet their Terms of Reference in terms of social learning and impact assessment. However, the Terms of Reference that call for but do not make sufficient provision for its undertaking – in terms of time, resources and team composition – are equally to blame.

There also remains the problem of the extent to which evaluators press on regardless to their conclusions and recommendations, demonstrating a lack of self-awareness among evaluators and an unwillingness to challenge the constraints placed upon them.

Spelling-out conceptually the different requirements for outputs, outcomes and impact analysis, as compared with problem identification and inputs, might help remedy this situation. Here again social, economic, cultural and political intelligence and outcomes and impact analysis are closely related.

3.4.9 **Attribution of Evaluator Bias**

If the concept 'independent' is difficult to pin down, 'bias' may be even more so – especially as this could be considered to be ever present in all social and other enquiry. The issue for meta-evaluation is how best to detect and demonstrate it, so that allowance may be made for it.

One of the co-authors of the 1999 DANIDA synthesis report, *'Evaluation of Danish Humanitarian Assistance 1992-98: Volume 1'* remarked on the extent to which he was able to trace 'evident bias'. Although the seven evaluations shared the same Terms of Reference and looked at aspects of the same donor programme (albeit in different countries), the six evaluation teams came from different institutions, with different histories of evaluation and composed of different sectoral bases. These characteristics shaped frames and discourses to the extent that if team names were covered up, anyone with knowledge of the evaluation community could have guessed which evaluator had carried out which evaluation.

However, the Kosovo evaluations did not have common Terms of Reference, and, though looking at the same conflict, were not all concerned with the same programme. It is therefore not immediately apparent how a heavily induction-led meta-

evaluation might identify independent variables or go about the rest of the task of attributing bias.

As for bias in respect of this meta-evaluation, there is no reason to believe it is any less influenced by its principal author's beliefs, experiences and formation as a social anthropologist. These include strong assumptions that commissioned consultancy and freely undertaken research deliver very different products, varieties of interdisciplinary scope, verifiability of findings and conclusions. This in addition to the view that evaluations of humanitarian action are best met by outcomes-oriented work, as is the case for evaluations of economic development issues.

3.4.10 **Conclusion**

The task of meta-evaluation is not easy, not always welcome, and often not understood. It is not easy because of, among other things, the frequently apples and pears nature of the evaluation reports reviewed. It is not always popular because evaluators and their managers may see it as a threat to the validity of not just exemplars of a genre but the genre itself.

Evaluation has been widely advocated as a path to lesson learning (as well as accountability), and despite the critique in this paper, the contribution to humanitarian knowledge and analysis that the Kosovo evaluations represent is itself a testament to what present practice can achieve.

The Kosovo reports are diverse. While some may be described as continuations of management by other means, others (some are mixed of course) suggest that evaluation might best serve lesson learning through advocacy, verification, and witness (particularly where protection is their main concern), rather than through would-be management mandates, business and organisational solutions.

A standard evaluation thesaurus sums up what are particularly evident as tools of inductive argument and critique as: 'analogies, examples, counter examples, counter explanations, more often than ... exact rules .. and the statements it use[s] ... are rough guides to the truth, that is, hints and heuristics rather than exact laws. We use certain modifiers – like *prima facies*, balance of proof, and *ceteris paribus* – sometimes *probably* – to flag the qualifications involved' (Scriven, 1991: p221).

Whether it would be well for different types of evaluation to proceed differently remains to be examined. With regard to lesson

learning, if the aim of evaluation is to do more than investigate mandates in order at least to meet the needs of results-based management concerns, it will have to find ways of becoming more outcomes-oriented than it is at present. This in turn means evaluation managers and evaluators becoming less socially blinkered or blind.

That more and more evaluations are being done and increasingly placed in the public domain – and given public consideration as in this publication – is, on any reading, excellent. What has not been explored in this chapter, however, is how the actual commissioning, designing and doing of evaluation could best be institutionalised in order for it to meet everything that is expected of it.

CHAPTER 4

OVERVIEW OF THE ANALYSIS AND NEXT STEPS

CHAPTER 4

OVERVIEW OF THE ANALYSIS AND NEXT STEPS

4.1 Introduction

This chapter combines the analysis of the principal findings and meta-evaluation contained in Chapter 2 and Chapter 3, thus covering 49 individual evaluation reports, five synthesis reports and one summary report.

In synthesising the principal findings from both the Kosovo and non-Kosovo sets, section 4.2 highlights the differences, considers likely contributing factors, and the implications of the overall findings for the humanitarian system and for ALNAP's own agenda.

As well as combining the meta-evaluation from the two preceding chapters, section 4.3 adds to the analysis of report quality through an application of the preliminary proforma to the Kosovo set. It concludes with a consideration of the next steps needed to strengthen the role of evaluation within the humanitarian system and, again, the implications for ALNAP.

4.1.1 Contextual Differences Between the Sets

The two sets of reports differ significantly in terms of the contexts in which the operations being evaluated took place. Of the complex emergency/natural disaster contexts covered by the non-Kosovo set only East Timor displays similarities to Kosovo, in respect of the efforts directed to peace enforcement, the setting up of an interim administration, nation-building, etc. Only one East Timor evaluation was available for inclusion in this Annual Review.

Detailed analysis of per capita expenditures in each of the operations evaluated was not readily available to the authors, however, perhaps the most striking difference between the Kosovo and non-Kosovo sets is the sheer scale of resources deployed in Kosovo, far outstripping those deployed in the other contexts (see 4.2.1 below).

Despite such obvious differences between the contexts, many of the Kosovo evaluations are comparable to the non-Kosovo evaluations in that they generally shared the same criteria, were

generally undertaken by consultants, and focused on one area or sector

4.2 **Principal Findings & Issues Raised by the Reports**

4.2.1 **Directions of funding**

The pronounced bias towards Kosovo in the allocation of funds by the humanitarian system has been raised by other publications (e.g., IASC, 2000), and discussed by one report as follows:

‘While donor governments gave US\$207 per person through the 1999 UN appeal for Kosovo, those in Sierra Leone received US\$16, and those in the Democratic Republic of Congo little over US\$8 ... The cost of Camp Hope in Albania for 3,000 refugees for two months was equivalent to the 1999 UN appeal for Angola ... In Africa for the past 20 years, refugees have often had to survive on lower than basic food requirements. Kosovo refugees in some cases were receiving Turkish delight in complimentary food packages. In the emergency, 80,000 refugees were flown to other countries in Europe, something that would never be conceived of in an African crisis’ (DEC, 2000c vol 1: p77).

Such disparities clearly show that the international humanitarian system is strongly influenced by factors such as national interest, geopolitical strategising and media focus. As a system it falls significantly short of meeting the criteria necessary for the use of the term ‘humanitarian’, such as impartiality, independence and considerations of proportionality, as contained in International Humanitarian Law and the Red Cross/NGO Code of Conduct.

A necessary first step in equipping the system to counter such biases, is the monitoring of humanitarian expenditure on a systematic basis. The Inter-Agency Standing Committee’s publication *Global Humanitarian Assistance 2000* (IASC, 2000) represents a welcome development in this regard. It was obliged, however, to make use of datasets that were not always comparable or consistent, and it is not clear whether it will be produced on a regular basis. Efforts need to be made by the DAC, OCHA and the system generally to strengthen the datasets essential to such monitoring.

A second step would be work to develop measures of humanitarian need to enable agencies involved in funding and coordinating a response, to determine at an earlier stage than at

present whether the humanitarian needs of affected populations (however defined) were being met, not met or significantly exceeded. Initial work in this field is planned by the Overseas Development Institute and Tufts University. Although it will inevitably encounter conceptual and technical difficulties, it can build on accepted existing approaches in the nutrition, food security, and public health sectors, and make full use of the widely recognised Sphere Minimum Standards in Disaster Response (Sphere, 2000).

Another funding related issue emerging from the reports is that a significant percentage of the funding being evaluated, perhaps between 30 and 50 per cent, concerned rehabilitation activities rather than activities which addressed 'acute human suffering'. These proportions are based on an impressionistic assessment of the activities evaluated and are very approximate. Nevertheless, they raise important questions as to the overall appropriateness of impact indicators such as 'number of lives saved' for an assessment of the overall intervention. In addition they point to the need for datasets on humanitarian expenditure to make a clearer differentiation between relief and rehabilitation.

4.2.2 **The Results of Humanitarian Action**

Judged on their own terms, interventions were assessed by the evaluations as being largely successful at meeting requirements and their stated objectives, with inputs leading to appropriate outputs. This in itself is a remarkable achievement. The efforts of agency personnel, which appear to be one of the most important factors contributing to success, need to be highly commended given the context and their difficult working conditions. At the same time, the claims of success need to be qualified.

As Chapter 2 and Chapter 3 point out, the stated objectives of interventions were often vague. The measures by which these objectives were assessed – mouths fed and lives saved – could have been more focused and disaggregated. In particular the Kosovo set tends to focus more on organisational issues and coordination or, at a lower level, on the results chain, either inputs or outputs. We learn for example that funds were used to rebuild housing, but not the uses of that housing (Tearfund, 2000).

Evaluation of impact – that is the real difference the activity has made (for a full definition see OECD-DAC 1999) – was required in many Terms of Reference, but commissioning agencies appear

to have been satisfied with reporting on effectiveness or input instead. The challenge in this case is promoting an understanding of the difference between effectiveness and impact, and ensuring that both areas are adequately addressed in evaluations.

Interventions were also assessed by the evaluations as being for the most part relevant to the affected populations in both the Kosovo and non-Kosovo set. In this case the standard of relevance does not appear to include systematic assessment of the views of the affected population, undermining overall findings.

In addition, the Kosovo and non-Kosovo evaluation reports were consistently weak in key methodological areas (notably triangulation, consultation with the affected population, and lack of attention to gender). They were for the most part opaque in methodology, to a much greater extent than is the norm in evaluations of development cooperation or the general evaluation field.

Few reports raised the issue of attribution adequately and only one attempted counter-factual analysis. As a result, the Kosovo reports in particular were unable to clearly isolate the main causes of results – e.g., between the good health and coping strategies of refugees, the support provided by host families and humanitarian action. Determining the factors responsible for a particular result is a major function of evaluation, and so this inability to isolate the main factors represents a significant shortcoming. This failure to establish credible indicators of success, combined with weak methodological approaches, must call into question the validity of the conclusions reached by many of the reports.

4.2 3 **Coverage**

The majority of evaluation reports concluded that coverage of affected populations and vulnerable groups within the programmes evaluated was at the least adequate, and that resources were not in general diverted. That resources were provided to vulnerable groups is a significant achievement and should help counter criticism in this area. On the other hand, both the Kosovo and non-Kosovo sets illustrate that the issue of coverage needs to be accorded more careful consideration by evaluations. In the Kosovo case a number of reports point out that the major focus of interventions in Albania and Macedonia was the refugee camp populations, to the exclusion of attention to refugees in host families. In the case of the non-Kosovo reports, the lack of

attention to IDPs is similarly noted. Another issue relating to coverage is that while assistance may for the most part be reaching vulnerable groups, this does not mean that all of the vulnerable population is being reached, a point not adequately considered in a majority of the reports

4.2.4 **Lack of Attention to Protection**

A common theme from the evaluation reports is that humanitarian agencies are not giving adequate attention to protection, humanitarian space and human rights in complex emergencies. This point comes out most clearly in the seven DANIDA evaluations which note to differing degrees the importance of protection as a key strategy. It is common for agencies to focus on the provision of material assistance and services such as health, food and shelter, leaving protection to agencies such as UNHCR and ICRC that have specific protection mandates. While UNHCR (2000a) notes the creditable performance of that agency in providing protection to refugees in the countries of asylum around Kosovo, Chapter 3 argues that lack of protection, and particularly armed protection of vulnerable minorities by the humanitarian system and the international community within Kosovo, should have been a main focus of those evaluations given the human rights violations there.

Within the humanitarian system there is now greater recognition than before that all humanitarian agencies have a role to play in relation to protection. A range of activities is underway which should contribute to an improvement in the situation. For instance the ICRC Protection Workshop Process, begun in 1996, has held four workshops so far (e.g., Von Flue and de Maio, 1999), and the IASC will shortly publish a practitioner's field practice manual (Paul, 2001).

However, debate continues as to what constitutes an appropriate role for different types of humanitarian agency. Paul (1999) notes that in considering their role in relation to protection, NGOs are often concerned with how their involvement in protection activities might affect their ability to carry out their mission; jeopardise their principle of neutrality; and risk their access to areas or countries controlled by groups/authorities antagonised by their protection activities. Addressing the issues raised by the evaluations is likely to require genuinely collective system-wide efforts to establish common understandings and approaches.

In addition, there is a parallel need to develop standards and

criteria for use in the evaluation of protection. For instance UNHCR (2000a) notes that UNHCR was criticised for being both too timid and too forceful in its relations with the Government of the Former Yugoslav Republic of Macedonia and in respect of the initial refusal of asylum to refugees. Evaluations can support the development of such standards, and the UNHCR evaluation (2000a: Chapter 6) is itself an example of good practice in the use of standards for assessment of protection (even if limited to refugees from Kosovo and not human rights violations in Kosovo).¹

4.2.5 **Coordination**

Coordination has a well-established reputation for being amongst the most problematic performance areas within the humanitarian system. That both the Kosovo and non-Kosovo evaluations support this view comes as no great surprise. Kosovo provides one of the most telling examples, where UNOCHA was not called on to play a main coordination role and UNHCR, designated as the UN Lead Agency, was to an extent marginalised by NATO and bilateral donors for what were largely political reasons (see Chapter 3 and UNHCR 2000a). There are also many other examples in the non-Kosovo evaluations (see Chapter 2).

Debates on the most appropriate modalities for coordination in the system continue, with little advance on those that began in the 1970s. While some praise and support OCHA, others criticise and undermine it. Likewise, some argue strongly in favour of the Lead Agency model, while others argue equally strongly against it. The adage that 'all agencies want coordination but none want to be coordinated' remains useful in explaining the coordination problem. However, the extraordinary persistence of the problem can probably only be explained in terms of structural factors such as the large number of separately-funded, separately-governed agencies typically involved in any single operation. Short of radical measures to reduce the number of agencies or channel funds through a limited number of central coordinating bodies, the system's structure will probably only achieve significantly improved coordination if it is able to develop an appropriate framework of incentives and disincentives for agencies to coordinate. Currently, the rewards for good coordination and the penalties for not coordinating are just too weak to have a noticeable effect.

The persistence of the problem also raises issues for evaluation. Are evaluations dealing adequately with coordination as a central

problem in the system? Almost certainly not, as their primary focus is on the performance of single agencies with coordination frequently regarded as a secondary issue. This argues for more collaborative evaluations. Are the necessary coordination standards and indicators in place for use by evaluators? Again the answer is negative, and this lack needs to be urgently addressed.

However, the picture provided by the evaluations on coordination is not all gloom. Examples of good practice are revealed in, for instance: OCHA's Mine Action programme in Afghanistan (DANIDA, 1999b); actors in the 1998 Bangladesh floods (DEC, 2000a); and the EMG in Albania (DFID, 2000). A common factor in each is a significant host country input – from Afghan staff in the Mine Action programme, to Bangladeshi government and NGO input, to the location of the EMG in the Prime Minister's Office in Albania. This relates to the need for good partners on the ground, a point agreed by all the reports that covered the issue. Partners were usually conceptualised in terms of INGOs, NGOs or multilateral institutions, but can clearly also include different levels of national government.

4.2.6 Linking Relief, Rehabilitation and Development

While the non-Kosovo set of evaluation reports raises LRRD as another of the most difficult areas, half of the Kosovo evaluations do not. Their focus is on the relatively brief period when the main refugee population was outside Kosovo. However, the seven ECHO Kosovo evaluations find that LRRD remains a major concern for ECHO, and the main points raised in the non-Kosovo reports are very similar to those in the Kosovo reports that do cover the issue:

- a lack of policy and strategy to guide LRRD;
- a lack of communication within and between agencies;
- an inability to plan adequate exit strategies.

These results are consistent with a major recent study on overall (rather than just humanitarian) assistance to post-conflict recovery:

'Despite ostensible good intentions, too often aid promised has not been committed, aid committed has not been delivered, and aid delivered has arrived too late. Moreover the planning and implementation of reconstruction aid has frequently suffered from inadequate preparation, poor coordination and lack of perseverance' (Forman and Patrick, 2000, p1).

The so-called 'grey-zone' between relief and development is not just a problem for the agencies being evaluated, it is also a problem for evaluation. Given that a significant proportion of humanitarian action budgets are being expended on both rehabilitation and development activities, evaluation reports need to be more specific in their recommendations concerning LRRD. An example of good practice in this regard is ECHO (2000p) that recommends a timeframe for hand-over; indication of the names of agencies intended to take-over work; and a mid-term review of the hand-over process. A useful exercise would be the collection and dissemination of good practice in the evaluation of LRRD by ALNAP. Ideally this would be undertaken in collaboration with the Conflict Prevention and Post-Conflict Reconstruction Network.²

4.2.7 **Coping Strategies**

There is evidence from the reports that the issue of coping strategies is on the agenda of some agencies, even if still dealt with in a sporadic fashion. While the Kosovo and non-Kosovo complex emergencies/natural disasters were very different events, affected populations used similar kinds of coping strategies. Primary among these strategies was mutual support. In the non-Kosovo cases Chapter 2 (Box 2.3) has already referred to the importance of neighbourhood solidarity in response to Hurricane Mitch, and formal and informal association in East Timor.

Most Kosovo evaluation reports note the reliance on host families and remittances from abroad, and some (e.g., ICRC, 2000) provide quite detailed information on feelings of mutual solidarity, as well as conflict, between host families and refugees. What the affected population was itself able to organise in the Kosovo case clearly outweighed the effects of humanitarian action, a point also made by the DEC with reference to natural disasters in Orissa and Central America (DEC, 2000, 2000b). In the Kosovo case, agencies failed to support adequately these coping strategies, partly because of their focus on refugee camps (see Chapter 3). A similar point can be made about the non-Kosovo interventions. A challenge for the future is to understand the different coping strategies of members of the affected population, and build on these during interventions.

4.2.8 **Knowledge of Context**

The evaluation reports demonstrate a major gap in agencies' understanding of the context in which they were intervening. In

Kosovo this was one of the most consistent messages from the reports, where the nature of 'ethnic' conflict in particular was inadequately considered. Without an adequate appreciation of context it is difficult, if not impossible, to design appropriate programmes. This is linked in turn to agencies' inadequate attention to coping strategies, where understanding how local and affected populations act in times of crisis is a major factor in understanding local context.

Why do agencies have such difficulty in gaining an understanding of the context in which they are operating? Among likely contributing factors is the lack of analytical tools for use by agencies, though the gap in relation to war economies is now being filled (Le Billon, 2000). Another factor is inadequate use of the capacities of existing local staff. However, the principal factor must be that agencies are simply not according sufficient priority to the development of the necessary analytical capacity, either within their own organisations or elsewhere in the system. Developing the capacity requires an investment that many agencies find difficult to justify or sustain when in direct competition with provision of assistance, or investments in what are perceived as more useful 'sharp end' activities such as logistics.

The reluctance of agencies working in challenging and dangerous environments to make such investments, may highlight the difficulties faced by a system composed of so many separate agencies. Some investments are seen to be uneconomical at individual agency level. The necessary economies of scale may be achieved by a number of agencies sharing costs. One example of such an initiative is the 'Learning Support Office' model (see Chapter 3, Endnote 5) being developed by ALNAP to provide services to a range of agencies – UN, Red Cross, NGO or others.

4.2.9 **Lack of Preparedness and Rapid Response Capacity**

Another consistent message from the reports is a lack of preparedness and capacity to respond expeditiously to sudden-onset humanitarian needs. In terms of preparedness, this was not just an issue of a lack of information but also a failure to act on available information. Preparedness does not appear to be an area bilateral donor organisations are overly willing to fund, preferring instead *ex post* interventions. This is perhaps due to the low profile nature of preparedness work and the possibility that funding preparedness may not have results.

Following their experience in Kosovo and elsewhere, several

agencies are in the process of setting up rapid response capacities with dedicated personnel or rosters of personnel deployable at short notice, linked in some cases to stockpiled materials. According to the ECHO evaluation (ECHO, 2000o) UNHCR is currently undertaking a programme which includes the following:

- enhanced systems of personnel recruitment and emergency deployment with 30 experienced personnel members, drawn from HQ and the field, on standby for emergency deployment within 24 to 72 hours,
- set up of an emergency pool of senior managers, to lead or coordinate large-scale and complex emergencies;
- intensive emergency training, especially at the senior and middle-management levels;
- external stand-by arrangements with a number of key partners strengthening emergency reserves;
- a database to determine the emergency capacity of NGOs.

However, agencies have made such investment in the past and, over time, the capacities have been eroded in periods of budget cuts, reorganisations or simply as a result of changed priorities. It remains to be seen whether the agencies will be able to sustain such investments this time around.

Once again the situation presents issues for evaluations. Standards as to what constituted an adequate level of preparedness in a given situation do not seem to have been available to evaluators, with the exception of fairly limited areas, for example, the need to have emergency teams on site within a certain period. The collection and dissemination of good practice case studies would be of assistance, and could be developed into guidance on how to evaluate preparedness.

4.2 10 **Inadequate Monitoring Systems**

The attention given by agencies to monitoring is found to be limited in almost all the evaluations. They also consistently lament the lack of systems that could provide data about results. Chapter 3 makes the point that an unmonitored programme is unevaluable in relation to standard performance measures.

Why monitoring is so weak is not clear, but may however relate to the difficulties of establishing effective systems in the first phase of response; the slow trickle down of performance measurement to humanitarian action; the idea that monitoring is not relevant in situations where life is in danger; and a lack of technical skills

of agency staff

In their monitoring recommendations, the reports often repeat evaluation mantras such as 'set clear objectives' and 'develop indicators', without recognising the constraints to such activities. DRC notes:

'With events moving so fast all the time in the Kosovo emergency it sometimes appeared pointless to those involved to use indicators of "success" and "failure" since the criteria for them kept changing. What at one stage appeared as a failure might later turn out as an achievement, or in any case as a learning experience' (DRC, 1999: p23).

Evaluators and evaluation guidance consistently underestimate the constraints to, for example, setting clear objectives – e.g., the need to negotiate with a range of actors; political constraints; pressures to allocate and disburse funds; and the need for specific technical skills. Unless constraints are recognised by evaluators and commissioning agencies alike, project managers are unlikely to pay adequate attention to general recommendations relating to monitoring and evaluation systems.

4.2.11 **Human Resources and Management Issues**

The picture that comes across from most of the reports is of under-trained, under-resourced staff doing their best in difficult conditions. Of all the problematics noted, this is one of the most important to rectify. Good performance crucially depends on the 'right woman or man, in the right place, at the right time', to paraphrase DFID (2000). The numerous recommendations on human resources fall into three main categories:

- the need for more senior better qualified personnel to be present at the beginning of the complex emergency/natural disaster;
- the need for personnel to be able to carry out adequate needs assessment including consultation with the affected population, gender and political analysis, and monitoring;
- less staff rotation.

One level of constraints relates to incentive structures within agencies. For example, a general lack of incentives to include gender, or to volunteer for field positions during complex emergencies (UNHCR, 2000a). A second level relates to lack of investment in personnel, including adequate training. In a culture of budget cuts and continuous organisational change, these