Joint Evaluations and Learning in Complex Emergencies

Lessons from the Humanitarian and Development Domains

Turid Lægreid

Joint Evaluations and Learning in Complex Emergencies

Lessons from the Humanitarian and Development Domains

Turid Lægreid

Introduction

With the growing number of actors and institutions involved in complex peacebuilding, there is also a growing demand for more unified and joint approaches to planning, implementation and learning. In UN peacebuilding missions, the *integrated approach* concept involves a mission design where the planning and coordination processes of the various elements of the UN family are integrated into a single country-level UN system.[1] NATO uses the term *comprehensive approach* for similar concept.

These approaches have been embraced by some, and strongly opposed by others. The first optimistic ambitions that military, political, development and humanitarian elements of all international intervention would become integrated in joint strategies, planning processes and implementation have been replaced by somewhat more realistic approaches. There has been gradual recognition that joint planning and implementation is highly complicated, and that the independence of the diverse organisations and institutions must be taken into account.[2] NATO and its members now refer more to the need for *coordination* of the diverse international actors than to *integration* of efforts.

This article argues that *joint evaluations* can be one tool useful for improving the coordination of very diverse actors. It examines some experiences of joint evaluations and learning processes undertaken mainly in the humanitarian domain, where valuable lessons have been learned over the past 15 years.

The international humanitarian community can be relevant as a point of departure, for several reasons:

- Humanitarian action is usually conducted in volatile, conflictive and unpredictable settings.

- Many diverse actors are involved, including donor countries, UN agencies, Red Cross/Red Crescent, NGOs, researchers, private contractors and recipient countries.

---

[1] Cedric de Coning, 'Implications of a Comprehensive or Integrated Approach for Training in United Nations and African Union Peace Operations.' NUPI, Oslo 2009.

- Cooperation and coordination mechanisms for policy development, planning and operational activities have been developed, and rest on the fundamental premise that actors and agencies are independent and their participation is entirely voluntary.

- The development of common policies and standards is self-regulatory, rather than imposed through any 'command' system.

This report examines the following specific cases: the joint donor evaluation of Rwanda (1995–96), the Tsunami Evaluation Coalition (TEC) 2005–2006 and the Inter-agency Real-time evaluation of Darfur (2004–2005).

These evaluations differed in their institutional frameworks and management arrangements, as well as to timing in relation to the outset of the crisis. On the other hand, they also shared some common features: they were all system-wide, as they included a broad range of actors and themes, and they all became references for joint evaluation policies and practices. Moreover, the evaluations dealt with crises that, because of the large-scale human suffering involved, had generated public interest and solidarity globally.

The report examines how challenges of management and organisation of the evaluations were solved, when such different and diverse independent actors set out to have joint learning processes. It looks at how the evaluations were initiated, what the purpose and objectives were and how they were managed. It will also see what main lessons and recommendations emerged from the evaluations, how these were followed up, and if any change or impact can be traced.
Some general lessons and experiences will then be extracted, in the hope that they may prove relevant also for future joint military and civilian learning exercises.

Why joint evaluations?

As defined by the OECD, an evaluation is

> an assessment, as systematic and objective as possible, of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.[3]

Evaluation is usually an inherent element in the planning and learning cycle – a good evaluation depends on there being well-defined goals and objectives in the planning process. It is intended to feed into further planning of projects and programmes, but also to support further policy development.

Joint evaluations have been used quite extensively in the development domain, particularly by bilateral donors. The OECD has indicated that the benefits of joint evaluations, as compared to single-agency evaluations, could include:[4]

- they provide mutual capacity development, as agencies and organisations can learn from each other, and share evaluation techniques;

- harmonisation – joint evaluations can limit the number of evaluation messages and foster consensus on upcoming priorities in a given country or setting;

- they have the potential to enable participation of developing country institutions (in the recipient country);

- they can increase objectivity and legitimacy – and not least – transparency;

- joint evaluations can address broader evaluation questions and can facilitate a perspective on multi-agency impacts beyond the results of one individual agency.

---

[3]   DAC Principles for Evaluation of Development Assistance. Development Assistance Committee. OECD, Paris, 1991,
[4]   OECD (2006), Guidance for Managing Joint Evaluations. DAC Evaluation Series.

The Joint Evaluation of the Emergency Assistance to Rwanda

The 1994 genocide in Rwanda shook the world. Within a few months, between 500,000 and 800,000 Tutsis and moderate Hutus were systematically massacred.[5] The international community response – or lack of response – was seriously criticised. The failure to prevent or halt the mass killings, despite the presence of a UN Peacekeeping Force, gave rise to questions about the capabilities of international conflict management. Also questioned were the performance and efficiency of the humanitarian agencies involved in the massive humanitarian operation in the aftermath of the genocide.

The idea of a system-wide evaluation was first launched by Danida in the OECD DAC setting, where it met some initial resistance. However, with the support of internal and external organisations, the idea gained momentum, and the process soon evolved to include a large number of different actors. In addition to the 19 OECD member countries, the core of the evaluation management consisted of representatives from EU, the DAC secretariat of the OECD, ICRC, IFRC, nine multilateral organisations (UN, IOM etc) and five NGO organisations/consortia.[6]

The objective of the evaluation was to:

> draw lessons from the Rwanda experience relevant for future complex emergencies as well as for current operations in Rwanda and the region, such as early warning and conflict management, the preparation for and provision of emergency assistance, and the transition from relief to rehabilitation and development.'.[7]

Four studies were commissioned: Study I – Historical Perspective, Study II: Early Warning and Conflict Management, Study III: Humanitarian Aid and Effects, and Study IV: Rebuilding Post-Genocide Rwanda.

---

[5] The exact number has never been established. This estimate was used by the Joint Evaluation, others argue that it was higher. See Human Rights Watch for a discussion of several estimates.
http://www.hrw.org/legacy/reports/1999/rwanda/Geno1-3-04.htm

[6] Australia, Austria, Belgium, Canada, Denmark, Finland, Germany, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the UK, the USA; the Commission of the EU, OECD/DAC secretariat, IOM, UN/DHA, UNDP, UNHCHR, UNHCR, UNICEF, WHP, WHO, IBRD, ICRC, IFRC, ICVA, Doctors of the World, INTERACTION, Steering Committee for Humanitarian Response, VOICE. (France suspended its participation in December 1995.)

[7] John Ericsson, 'The International Response to Conflict and Genocide: Lessons from the Rwanda Experience. Synthesis Report'. Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda. Copenhagen, 1996.

Evaluation management

A thorough management structure was set in place. A Steering Committee, made up of 'interested members of the international aid community', met four times.[8] Its main tasks were to finalise the Terms of Reference (ToR) and approve the short list of evaluators, approve budget and funding and discuss and provide feedback on study reports, approve outlines and later the draft of the synthesis report. The Steering Committee was also in charge of presenting the final report to the international community.

In addition, each study was managed by a lead donor agency (I Sweden, II Norway, III UK, IV USA). These four lead agencies, with Denmark in the chair, constituted a management group that was to contract, assign and supervise the work of the consultants and institutions in charge of the studies. In addition to the formal management structure, the evaluation also made use of consultations with experts. A group of resource persons, selected because of their expertise and experience, was invited to read and comment on presentations and drafts.

The original intention had been to engage Africans in the evaluation. However, no representatives from African civil society or governmental institutions were included in the formal management structures. Instead, a panel of seven experts from Africa provided a critique of the report through participation in two panel discussions with the authors of the reports and selected resource persons.

Implementation and scope

The scale of the evaluation process was unprecedented. In all, 52 researchers and consultants were employed on the five studies, and the cost of the whole process was USD 1.7 million. Study 3 – on humanitarian aid – cost USD 580,000 and had a team of 20 specialists and support staff with a combined input of four person-years.

The evaluation teams had several field visits to Rwanda and the African Great Lakes Region, as well as visits to UN Headquarters in New York and Washington DC, Paris and Brussels.

To ensure cross-fertilization and joint analysis, several workshops were held in the course of the year that the evaluation lasted, where the various evaluation teams met and discussed preliminary findings and analytical approaches.

---

8   John Borton and John Eriksson, 'Lessons from Rwanda – Lessons for Today: Assessment of the Impact and Influence of the Joint Evaluation of Emergency Assistance to Rwanda.' Danish Ministry of Foreign Affairs (2004). The International Response to Conflict and Genocide: Lessons from the Rwanda Experience.

Given the sensitive nature of the crisis, the evaluators did encounter some resistance when undertaking data collection and when presenting their preliminary findings. UNDPKO was initially very reluctant to acknowledge that they had received early warning on the genocide from the Force Commander of UNAMIR, the peacekeeping force, and did not allow access to the files.[9] UNHCR became involved in a dispute with the evaluation team (Study II) on the findings and conclusions on its performance. The French Ministry of Defence was reluctant to disclose its policies and activities in Rwanda, and France later withdrew from the steering committee of the evaluation because of the criticism of the French involvement.

Main lessons emerging from the evaluation
The evaluation listed 24 main findings, each with a set of corresponding recommendations. According to the Former Chairman of the Steering Committee for the evaluation, these findings can be summarized in the following main lessons:[10]

1. Unwillingness to acknowledge that genocide was happening prevented action to stop it.
2.   In view of the scale of suffering, the humanitarian aid was impressive, but could have been better prepared, better coordinated and more cost-effective.
3. Inadequate and slow support for reconstruction of the Rwandese society and government, including the justice system, has delayed a return to normalcy.

The evaluation highlighted severe systemic weaknesses in the international community, with a prime focus on *lack of policy coherence.* The evaluation pointed to the fact that policy and strategy formulation had failed to take political/diplomatic, humanitarian, military/peacekeeping and development aspects into account in an integrated matter. The conflicting interests among Security Council members, the discrepancies between the UN Secretariat and the Security Council, as well as inadequate communication within the various functions of the UN Secretariat all contributed to this lack of coherence.

The evaluation also recommended that the main international actors, including the main bodies of the UN, the OAU and the member states, should take measures to effectively prevent and suppress genocide and to include protection of civilians in peacekeeping operations.

On the humanitarian side, the evaluation stressed that responsibility

---

[9] The author supported the Study II evaluation team, and participated in the data collection in New York and Paris.
[10] Nils Dabelstein, preface in Borton and Eriksson (2004).

for inadequate systems for contingency planning and preparedness measures had to be shared between donors and operational agencies, and it offered several recommendations to increase funding and concrete measures for improving preparedness. The performance of the numerous NGOs involved was found to be of mixed quality. While some acted with professionalism and delivered high-quality assistance, others acted unprofessionally and irresponsibly.

Follow-up of implementation

The steering group decided to have a follow-up review of the recommendations after six months. At its meeting in November 1995, the Steering Committee agreed to review the impact of the JEEAR reports one year after their publication; and a second process, the Joint Evaluation Follow-up, Monitoring and Facilitation Network (JEFF), was set up to monitor and report on the evaluation's 64 recommendations. JEFF was a small network of 11 individuals representing the Management Group, the study teams and the Steering Committee, with a part-time secretariat and a modest budget. In the 15 months following publication, JEFF members participated in a total of 73 events. The final JEFF report was issued in June 1997, 15 months after the publication of the evaluation itself.

Two-thirds of the recommendations were judged to have had at least some positive outcomes. Main areas of progress were found to be: the strengthening of human rights machinery in Rwanda; the development of early-warning information systems in the Great Lakes region; the broadly supported efforts within the NGO community to improve performance through the development of standards and self-regulation mechanisms; and? the commitment shown by donors, UN agencies and NGOs to improve accountability within humanitarian aid.[11]

However, this first review found that no progress could be detected for the following recommendations:

- Fostering Policy Coherence (directed at the UN Security Council, Secretariat and General Assembly)

- Effective Prevention and Early Suppression of Genocide (directed at the UN Security Council, the Secretaries-General of the UN and Organisation of African Unity (OAU) and the High Commissioner for Human Rights.

---

[11] John Borton, 'The Joint Evaluation of Emergency Assistance to Rwanda', *Humanitarian Practice Network,* 26, March 2004.

Danida, the Danish development agency that had taken the initiative to the evaluation in 1995, commissioned a study 10 years after, to assess the influence and impact of the evaluation.[12] The study revealed that the Joint Evaluation had been highly influential: it had stimulated further research and analysis of complex emergencies and the tendency toward genocide, and had been used widely as a teaching resource at the college and university levels. It was seen as pioneering, as it was the first report to be published that assessed the performance of the international community in all phases of the genocide, and because of its unusually wide scope. A further factor was the inclusive governance of the Joint Evaluation, which gave political weight to the evaluation. The fact that its Steering Committee included representations of the same countries and institutions as those to whom the evaluation was addressed, increased its credibility.

The most direct impact had been in the area of humanitarian accountability. The development of the 'Sphere Standards' was a benchmark in creating common standards in emergency settings. Although the process had been initiated before the evaluation, it is widely recognised that the evaluation gave the then fragile and informal process momentum and legitimacy, and helped move the process forward.[13]

Other more or less direct spin-offs were the Humanitarian Accountability Project (HAP) and the creation of ALNAP – The Active Learning Network for Accountability and Performance in Humanitarian Action. ALNAP has become a unique network that incorporates most of the key humanitarian organisations and experts, including members from donors, NGOs, the Red Cross/Red Crescent, the UN, independent experts and research institutions.[14]

Since the first review of recommendations in 1996, there have been important policy changes within the international community concerning prevention of genocide, some of which were only partially captured by the 2005 assessment. The same year, world leaders agreed through the World Summit Outcome Document that each individual state has 'the responsibility to protect its populations from genocide,

---

12   John Borton and John Eriksson (2004). The Terms of Reference set the purpose of the assessment to be four-fold. Using the Joint Evaluation as a basis, the assessment was to: 1) assess the extent to which the evaluation has been acknowledged in subsequent literature and official reports, and by key informants; 2) review developments in the fields/sectors covered by the Joint Evaluation recommendations since publication in 1996; 3) ascertain the degree to which developments in the different sectors that the Joint Evaluation addressed are consistent or inconsistent with Joint Evaluation recommendations.

13   See Margie Buchanan-Smith, *How the Sphere Project Came into Being: A Case Study of Policy-Making in the Humanitarian Aid Sector and the Relative Influence of Research.* Working Paper 215. Overseas Development Institute, London 2003.

14   See www.alnap.org for more information on this network.

war crimes, ethnic cleansing and crimes against humanity' and that this entails the prevention of such crimes.[15]

However, little progress has been made when it comes to actually preventing or halting genocide. The Darfur tragedy is an example of a situation where information was available, political and public attention was very high, but where again the international community failed to prevent mass killings and provide protection for the civilian population.

The Tsunami Evaluation Coalition (TEC): Joint evaluation of the international response to the Indian Ocean tsunami

The Rwanda joint evaluation has been a reference point for joint evaluations of humanitarian response. However, despite the influence and impact it had on the humanitarian community, and the growing interest for learning and accountability in the humanitarian sector, it took more than 10 years before another large-scale joint evaluation of a humanitarian crisis was undertaken.

Only a week after the December 2004 tsunami, several humanitarian agencies began to discuss how to coordinate evaluations of the tsunami response, and in February 2005 a consultative inter-agency and donor meeting was convened. The agencies were interested in how to maximise learning through joint evaluations. Five thematic areas were identified for further investigation: coordination, needs assessment, the impact of the international response on local and national capacities, links between relief, rehabilitation and longer-term development, and the funding response.[16]

Management
A Core Management Group (CMG), consisting of 14 representatives from humanitarian and development cooperation agencies, managed the strategy and process, through general oversight and direction for the TEC on behalf of its wider membership. The ALNAP secretariat provided the platform. For three of the thematic evaluations, multiagency Steering Committees were formed to take decisions on such matters as the selection of the evaluation teams, as well as providing guidance and advice to the evaluation teams, and signing-off on the

---

[15] UN General Assembly A/RES/60/1 "2005 World Summit Outcome". http://daccessdds.un.org/doc/UNDOC/GEN/N05/487/60/PDF/N0548760.pdf?OpenElement
[16] J. Telford and J Cosgrave (2006): Joint Evaluation of the international response to the Indian Ocean tsunami: Synthesis Report. London: Tsunami Evaluation Coalition.

final report. As the Steering Committees included operational agencies, the members were also useful in ensuring that field representatives were aware of the TEC, and available to contribute to the evaluations.

Implementation and scope

The TEC was the largest humanitarian evaluation ever undertaken. The total cost was approximately USD 2.9 million, and some 30 independent consultants and researchers were contracted for the main evaluation teams. The teams conducted field visits of two to three weeks to the most affected countries (Indonesia, Sri Lanka, Thailand), and more than 40 persons participated in additional research studies. [17]

Main findings

One of the main findings of the evaluation concerned the failure of international agencies to build on local capacities, even though the immediate life-saving support had consistently been provided by neighbours, local authorities and other local actors. The evaluation also pointed out that the international agencies to a little degree coordinated their response, instead competing for funding in one of the best-funded (if not over-funded) crisis responses ever.

The authors of the synthesis report noted that these are systemic weaknesses of the humanitarian response system, first identified in the Rwanda evaluation.

The four main recommendations from the evaluation were the following:

1. The international humanitarian community needs a fundamental reorientation from supplying aid to supporting and facilitating communities´ own relief and recovery priorities;

2. All actors should strive to increase their disaster response capacities and to improve the linkages and coherence (...) in the international disaster response system, including those from the affected countries themselves;

3. The international relief system should establish an accreditation and certification system to distinguish agencies that work to a professional standard in a particular sector; and

---

[17] For the funding study alone, more than 30 sub-studies were commissioned for each of the main donor countries and main agencies.

4. All actors need to make the current funding system impartial, and more efficient, flexible, transparent and better aligned with principles of good donorship.[18]

Follow-up

In the immediate aftermath of the evaluation, the TEC findings and conclusions were extensively discussed in workshops, meetings and seminars, mainly within the coordination structures and networks (IASC and ALNAP) in which the TEC had originated. The main agencies forming the TEC commissioned a follow-up evaluation three years after the tsunami, as the conclusions of the reports had indicated serious concern that the developmental aims of the response were not likely to be achieved.[19] The follow-up focused specifically on the linkage of relief, rehabilitation and development; and while the first evaluation had dealt mainly with process and policy issues, the time elapsed since the initial response allowed the follow-up evaluation to conclude on results and achievements as well. A new element in the management structure was the inclusion of the government agencies of the affected countries.[20]

---

[18] Ibid

[19] 'A ripple in development? Long-term perspectives on the response to the Indian Ocean tsunami 2004. A joint follow-up evaluation of the links between relief, rehabilitation and development (LRRD)'. Sida 2009.

[20] Joint Steering Committee: Sida, Norad, Danida, the Netherlands Ministry for Foreign Affairs, CIDA, BAPPENAS Indonesia, BRR Indonesia, Ministry for Plan Implementation, Sri Lanka. Ministry for National Building, Sri Lanka, ISDR, Bangkok, IFRC Bangkok, CARE International, OCHA and UNICEF.

Inter-Agency Real-Time Evaluations

Real-time evaluations have become increasingly used tools for humanitarian agencies. The UNHCR has been the agency that has most consistently made use of such evaluations, and has also developed tools and methods that have since been adopted by other agencies.

The recently published guide to real-time evaluations[21] defines a real-time evaluation as 'an evaluation in which the primary objective is to provide feedback in a participatory way in real time (i.e. during the evaluation fieldwork) to those executing and managing the humanitarian response'. Real-time evaluations differ from 'ordinary' evaluations not only in timing and speed, but more importantly in design, that they are meant to be interactive, and that they are more directed to operational managers in the field than the case with other evaluations.

Inter-agency real-time evaluations have begun to be tested out only recently. The Darfur real-time evaluation was one of the first, and was a valuable learning exercise for what might become a standard feature of humanitarian response. There is a growing recognition that timely evaluative feedback into ongoing humanitarian operations on issues that go beyond single-agency programmes is useful for improving system-wide performance. This can offer a unique framework for inter-agency system-wide evaluation by reviewing the overall direction, coordination and implementation of an emergency response, rather than the solely agency-specific aspects, thus offering a view of the broader effects at the level of the humanitarian system. Since the Darfur Real-Time Evaluation (RTE) in 2004/2005, inter-agency RTEs have been used in the humanitarian response to the Mozambique floods, the Pakistan earthquake response (with a special focus on the cluster approach) and the drought response in the Horn of Africa. Most recently, an RTE was launched to give feedback on the response to the Cyclone Nargis (Myanmar).

The Darfur Real-Time Evaluation

The inter-agency real-time evaluation of the Darfur crisis (Darfur RTE) was commissioned by the Emergency Relief Coordinator in August 2004, one year into the crisis, to look into concerns about the

---

[21] John Cosgrave, Ben Ramalingan and Tony Beck (2009) 'Real-time evaluations of humanitarian action. An ALNAP Guide. ODI 2009 (Pilot version). *www.alnap.org/publications/pdfs/RTEguide.pdf*

timeliness and effectiveness of relief delivery to the population of Darfur, Sudan.

The background was the slow international response to the crisis. It had become apparent that huge humanitarian needs were still unmet, and the civilian population virtually without any protection of their basic rights.[22]

The objective of the evaluation was twofold: to recommend actions that might be taken to improve the operational response in real time; and to identify broader lessons learned in Darfur for future humanitarian action there and elsewhere.[23]

Four main issues were to be examined:

- – leadership and coordination

- – overall operational response for the UN agencies, OCHA and NGOs

- – advocacy work of the UN and OCHA

- – donor response.

The evaluation was to focus on the UN response, but would also examine the role and performance of NGOs and donors.

Management
The overall management of the evaluation was left to one lead agency – OCHA. Its tasks included responsibility for the conduct of the evaluation, assessing the quality of the reports, day-to-day contact with the consultants and keeping stakeholders informed. The interagency aspect was taken care of by establishing a 'core learning group' involving key operational IASC members, where both evaluation staff and programme staff participated.

Implementation and scope
The evaluation was conducted by two independent consultants. To ensure learning, the team was also supported by staff from OCHA and Care who were not directly involved in the operational response. The real-time evaluation was designed in three phases, with three field visits, over a period of ten months. Each visit was followed up with sepa-

---

[22] The humanitarian response later developed into one of the most massive humanitarian operations ever.
[23] The evaluation was to examine the timeliness, coherence, coverage, appropriateness, relevance, efficiency and effectiveness of humanitarian and protection provided to the vulnerable population.

rate reports.[24] The purpose of iterative visits was to enable the team to observe responses to the crisis as it unfolded, and feed suggestions for immediate course corrections into existing mechanisms and fora.[25]

Information was collected through interviews with key stakeholders, focus groups, surveys, observation and review of background documents. Participatory workshops were conducted to ensure interactivity.

Findings and recommendations
A main finding was that nearly all the agencies and organisations involved were unable to mobilise the appropriate capacity to respond to the Darfur crisis. There were also deficiencies in leadership and coordination that could have been improved with deeper UN field presence, and clearer distribution of roles and responsibilities for critical areas. Also highlighted was the lack of accountability for performance, for both NGOs and UN agencies – this was linked to the absence of agreed benchmarks. The evaluation also concluded that an agreed framework to protect human rights would have been useful, and that more, not less, public advocacy on behalf of the war-affected populations would have helped to create a more protective environment.

In particular, the RTE recommended improving the accountability of headquarters support to the field, extending the UN presence beyond state capitals, strengthening sectoral planning at the field level, and installing a strategic planning and analysis unit. The evaluation also recommended that a common advocacy platform for the UN be developed.

Follow-up of recommendations
To facilitate follow-up of the recommendations, a management response matrix was developed. Each of the specific recommendations was broken down into action points, with defined responsible focal points.

A first lessons-learned workshop was conducted in the Inter-Agency Standing Committee's Working Group in March 2006, and OCHA has since attempted to track the follow-up of recommendations. The conclusion is that there is mixed record of follow-up. While the Darfur RTE worked as a catalyst for action on several of the issues, OCHA has recognised that most of the recommendations were not acted upon in real time. As reasons, they note competing priorities, the sheer

---

[24] A fourth and final report was planned, but was never completed due to illness. The main report thus consists of the three sub-reports.
[25] Bernard Broughton and Sarah Maguire: 'Inter-agency Real-Time Evaluation of the Humanitarian Response to the Darfur Crisis'. OCHA 2006.

number of recommendations, a lack of capacity at the field level, and institutional insularity.[26]

Here it should be added that several of the recommendations reflect systemic weaknesses in the humanitarian community, and are thus less likely to be acted upon promptly by operational staff. Issues like humanitarian leadership, coordination, planning and accountability – all central to the findings in the RTE – are main elements of the Humanitarian Reform Process launched in 2005 by the Emergency Relief Coordinator. This reform was mainly based on a Humanitarian Response review that went parallel to the initial Darfur response. The weaknesses found in the review were also noted by the Darfur RTE. In that sense, many of the recommendations are indirectly being dealt with through the Humanitarian Reform, as various specific measures have been put in place to improve humanitarian leadership, cluster planning (or sector planning), predictable distribution of responsibilities for response and accountability measures. However, the improvements are most evident at policy level, and the reform is still in an early implementation phase.[27]

---

[26] OCHA Evaluation and Studies Section (ESS), 2006 Report.
[27] See http://www.humanitarianreform.org/ for more information on the elements of the reform, and status of implementation.

Lessons from Joint and Inter-Agency evaluations

With the increasing interest in and use of joint and inter-agency evaluations has also come a growing realisation that these are complex undertakings, and that more knowledge and lessons need to be drawn from the experience gained. Within the ALNAP network, several workshops and sessions have sought to summarise lessons learnt so far, and the IASC has established a specific subgroup for inter-agency real-time evaluations. In the following we present some general lessons, drawn partly from these lessons-learnt exercises, the cases studied above, and from the general literature on joint evaluations.

The literature and guidance notes developed to date have been drawn mostly from bilateral development support in stable and traditional long-term donor–recipient relations. Such generic lessons can nevertheless be of use, as the benefits and challenges are often similar. However, it should be noted that the challenges will multiply with the number and diversity of the actors; moreover, the interventions and programmes to be evaluated will often have been conducted in conflict settings marked by considerable unpredictability.

The OECD guidelines for managing joint evaluations highlight the need to establish, at an earlier stage, common ground rules for the evaluation: 'It should not be assumed that each of the partners has the same objective for the joint evaluation. Experience shows that conflicting – and sometimes hidden – agendas often emerge as the evaluation process develops.'[28] Agreeing on common ground rules includes deciding on who should be involved and how, what purpose and objectives the evaluation should have, making sure that all stakeholders share the main principles and standards for the evaluation, and deciding on sound management structures.

Although the evaluations studied involved a wide range of organisations and institutions, there were no reports of big discrepancies. The stakeholders and the evaluators in the Rwanda evaluation all agreed that the performance of some NGOs had been poor, that lives had been lost as a result, and that something had to be done to raise the quality of humanitarian assistance. In retrospect, the main evaluators

---

28 OECD (2006).

identify this fact as a success factor for the evaluation itself, as well as for the impact it was later to have on the humanitarian system.[29]

The fact that the stakeholders in the Darfur RTE and the Tsunami Evaluation shared a common general understanding of what the main important issues and the problem areas were facilitated the conduct of these evaluations. This does not mean that there was any lack of discussion and exchange of views, but rather that the parameters for analysis were shared, and the standards for judging success, etc., broadly the same.

The humanitarian community has come far in harmonising standards and overall policies for humanitarian action. Regular meeting points and coordination structures have facilitated this. In part, the Rwanda evaluation was a main factor in the creation of learning networks like ALNAP. This network, and the harmonisation of humanitarian tools, principles and standards it has promoted, in turn served to facilitate the Darfur evaluation and the TEC.

It should be noted that all of these evaluations, as is the case with most evaluations of development and humanitarian work, were based on the OECD DAC Principles for Evaluation of Development Assistance.[30]

Agreeing on who should be involved
Early on in the evaluation, there should be a decision on who should be involved. A broad, inclusive approach can create more ownership to the evaluation and its findings – and can lead to a higher degree of follow-up of recommendations. However, if too many actors are involved, that can lead to management problems and impede progress with the evaluation. Additionally, the transaction costs involved in joint evaluations can be particularly challenging for smaller organisations.

The three evaluations studied here were initiated in different ways. The Rwanda evaluation came about at the instigation of bilateral donors within the formal OECD DAC framework, with Danida as the primus motor, but it soon involved a range of actors. The Darfur RTE was initiated within the Inter-Agency Standing Committee, also with a set membership but including various actors, and with OCHA as the main driving force behind the evaluation. And the TEC had a much more 'democratic' start, with 'interested agencies and institutions', all connected within the informal network of ALNAP, setting off the

---

29 Borton (2004).
30 OECD (1991). Humanitarian agencies also make use of the ALNAP initiated guide 'Evaluating Humanitarian Action using the OECD-DAC Criteria' by Tony Beck. Final version? March 2008.
      http://www.alnap.org/resources/guides/evaluation/ehadac.aspx

evaluation. However, there is little to indicate that whether the initiative was taken within formal or informal structures has had much impact on the evaluation itself.

The OECD guidance note recommends that

> ground rules should clarify the roles and responsibilities of each partner and establish clear mechanisms for decision making and cost/burden sharing. The evaluation ground rules should ensure that the views of the less powerful partners will be respected and should define the commitment that all partners will have to the findings and recommendations.

The lessons-learnt exercise held in the IASC setting after the Darfur Real-Time Evaluation indicated that not all stakeholders had shared an understanding of the value of the evaluation, and the institutional commitment to the evaluation varied. It was also indicated that the NGOs could have been more involved, to ensure that the application and results of the evaluation would benefit not only the UN system but NGOs as well.

The TEC was assessed as having been quite successful in involving the main stakeholders. There were differing layers of involvement, which allowed smaller organisations to be involved in different ways. That is a factor to take into account when more diverse actors plan to undertake common evaluations – not least in politically controversial settings. Diversity in itself is not necessarily a problem, as long as there are formal and informal meeting points, shared standards and principles, and a shared understanding of the strengths and weaknesses of the responses in general.

Equally important is deciding *how* the various stakeholders should be involved: who should be in the lead, who should be deeply involved in most steps of the evaluation, and who should generally be kept informed only about major decisions and developments.

Field-level involvement
Most joint evaluations in the development and humanitarian domains have been led by the headquarters of the organisations in question, and mainly by their evaluation and monitoring managers and officials. When the chief aim is to learn how to improve future policy, a main concern will usually be to ensure that top management and policy departments in the various organisations are involved and interested, as they will usually be the ones that can ensure follow-up and implementation of recommendations.

With RTEs, however, the main target audience should be operational staff and managers – primarily at field level, but also at head offices. Communications between head offices and the field proved difficult for the Darfur RTE. Several field managers complained that the RTE was a top–down affair, and that the main purpose was accountability from the field to the head offices, and not learning.[31] How to include the field level remains a major challenge for all RTEs, particularly inter-agency ones.

Evaluations also involve a burden on the field level, which will need to balance the priority of ongoing operational activities with dealing with evaluation teams and their need for support. The Tsunami Evaluation was not intended to have impact and influence in real time, but it was started during the first phase of the response, and field visits and interviews were conducted while operations were ongoing. The TEC initially struggled with the funding aspect, which delayed the contracting of evaluation team leaders, resulting in insufficient lead-time between team formation and field visits.[32]

Some resistance was also reported in connection with this evaluation. It was held that some senior agency managers (particularly within the UN) were not interested in sharing information for evaluations that had not been commissioned directly by the agency itself or their donors.[33]

Host-country involvement
Ideally, the host country and the receiving community should be equal partners in evaluations. Indeed, for 'traditional' joint donor evaluations in developmental contexts, host-country involvement is a main reason for conducting joint evaluations, as it will provide better oversight of aid impact for the receiving country.

In conflict and/or humanitarian settings, deciding to involve the host country needs thorough discussion and deliberation. If the host government is a party to the conflict – as was the case with the Sudanese government in Darfur – direct involvement would easily be counterproductive.

The Rwanda evaluation and the TEC were carried out in less conflict-filled settings, where the host countries could have been involved. In the evaluation reports and follow-up meetings, the evaluation manag-

---

31 Peta Sandison, 'The utilisation of evaluations', Chapter 3 in *ALNAP Review of Humanitarian Action in 2005. The utilizations of evaluations.* ALNAP, 2006.
32 TEC, 'Lessons Learned from TEC Process', report from TEC Review Day 14 February 2006, London. From http://www.tsunami-evaluation.org/NR/rdonlyres/9DBB5423-E2EF-43AB-B6D2-2F5237342949/0/tec_lessonslearned_ver2_march06_final.pdf
33 Telford and Cosgrave (2006)

ers took self-criticism for the failure to involve local actors systemati-
cally in the evaluation processes. The TEC partly made up for this
flaw by including the affected governments in the management struc-
ture of the later follow-up evaluation.

Deciding on the main evaluation issues to explore

The lessons-learnt processes after the Darfur Real Time evaluation
and the Tsunami Joint Evaluation concluded that both these evalua-
tions could have benefitted from more thorough discussion of which
main issues to explore. Many of the stakeholders involved the Darfur
RET commented that the ToR were too broad, and included too many
diverse issues.[34] Having a large number of issues will inevitably lead
to a large number of recommendations, which in turn can disperse re-
sponsibilities for follow-up action. For the tsunami evaluation, the
number of issues was not a problem, but the stakeholders reflected
later that the issues had been decided somewhat at random, and that a
more thorough discussion would have advantageous.

The OECD guidance note on joint evaluations stresses the importance
of holding early discussions while openly acknowledging the motives
and incentives of the different partners and anticipating possible prob-
lems.[35] For broad evaluations, it recommends that an initial scoping,
options or approach paper be commissioned. This should explore the
main issues, challenges and opportunities and provide the basis for an
informed discussion between all the partners on the overall approach
to the evaluation.[36] Civilian and military actors will often differ in
their approaches to evaluations and learning, with different methods
and terminology.[37] Such a concept note or approach paper could be
particularly useful in providing common grounds for the further proc-
ess. If possible, it should also attempt to differentiate between diver-
gent interests and analysis on one hand, and differences in language
and terminology on the other.

**Matching ambitions (purpose and objectives) with resources, de-
sign and scope**

The three evaluations differed in their purpose and scope. While the
Rwanda and TEC had system-wide learning as the main purpose, the
Darfur Real-Time Evaluation was intended first and foremost to en-
sure immediate improvements in field performance.

---

34 Sandison (2006).
35 OECD (2006)
36 Ibid.
37 Kristin M.Haugevik and Benjamin de Carvalho: 'Civil-Military Cooperation in
    Multinational and Interagency Operations', Discussion Paper on Operational
    Terminologies and Assessment for Multinational Experiment 5 (MNE5). 718
    Working Paper, NUPI, Oslo 2007.

Moreover, the resources made available for the Rwanda and TEC, in terms of person-days spent and budgets, were unprecedented. With several million dollars in funding, and some 40 to 50 consultants and researchers, they were able to provide in-depth studies on a range of issues, and succeeded in providing well-founded recommendations on overall policy issues. The inclusion of the main donors in the evaluation management was without doubt a contributing factor to the scope of these evaluations. The financial question was by no means an easy one, as delays in the provision of financial contributions to the TEC delayed the formation of the evaluation teams and thus the fieldwork of the teams.

In contrast, the Darfur Real Time evaluation had a small team of two external consultants and a correspondingly small budget. Although that was a necessity to ensure fast deployment and a lean process, the issues to be covered were quite broad, ranging from overall policy issues to fairly technical programmatic problems. That might help to explain why there came to be rather poor follow-up of the many recommendations resulting from the evaluation.

The tendency to overload the evaluation with issues and objectives is prominent also in single-agency evaluations, as various sections and actors within the organisation may have differing interests and perspectives on a programme, and will attempt to throw in questions from their own agenda. With joint evaluations, this tendency will only be magnified. A rigorous process, at an early stage, of narrowing down the main focus, purpose and objective will facilitate the later work. A realistic plan for the financing, including the burden sharing among the participants, should be part of the initial process.

Accountability or learning?
Most evaluations have elements of both learning and accountability. However, it is important to establish what the main purpose is, as that will influence further evaluation choices, and not least the various actors´ follow-up of the evaluation.

Accountability is about transparency in how resources have been used, and what results and (possibly) impacts have been achieved. Whereas accountability has more to do with the past, learning needs to be more forward-looking, aimed at producing some sort of change – at programme level, organisational or institutional level, even at system level.

In a survey of participating agencies in the TEC, the majority cited 'learning' (for improved performance) as their main rationale for tak-

ing part in the joint evaluation, while accountability came fifth.[38] Most humanitarian agencies will also need to have a very detailed evaluation of their own programme outputs and results, for purposes of accountability. A system-wide evaluation will usually not provide such level of detail.

For joint evaluations involving both military and civilian actors, it would be particularly advantageous to put the emphasis on learning. The accountability element should focus on transparency to the affected communities and states, seeking to explain what has been achieved by the many foreign interventions, rather than to home constituencies – who are usually more concerned about how effectively taxpayers´ money has been used.

Agreeing on a clear management structure
The Rwanda evaluation was a very ambitious undertaking, in terms of the number of issues, the controversy surrounding the crisis, and the number of stakeholders and consultants involved. The clear, transparent and inclusive management structure has been highlighted as one of the main factors that enabled the report to be finalised in the course of one year. The Rwanda evaluation model of having one small management group with a broader steering committee was repeated for the tsunami evaluation, which was also assessed as having worked very well.

The Rwanda evaluation and the tsunami evaluation are comparable in management structure and in process design. The decision to divide the main tsunami evaluation into sub-themes, with separate evaluation teams and corresponding steering groups, proved to work well in the view of most stakeholders, including the evaluators.[39] The committees that were set up to follow each of the sub-evaluations provided direction to the evaluation teams and functioned as quality assurance, in addition to serving as a link to the field organisations.

If an evaluation is divided into sub-studies, meeting points should be created between the various evaluation teams to ensure cross-fertilization. This was seen as a strength of both the Rwanda evaluation teams and the TEC teams, who were able to meet to discuss with each other, as well as with policy-makers, evaluation managers and independent experts.

Another important lesson for the system-wide evaluation is the need to have full-time secretariats to coordinate research, information flows

---

38 Report on joint evaluations, ALNAP
39 ALNAP Workshop report on TEC (ALNAP Biannual meeting, Rome, December 2006)

and administrative issues. For the TEC, the decision to delegate these functions to the ALNAP secretariat was assessed to be a success factor for implementation of the evaluation.

Such elaborate structures would not be suitable for real-time evaluations, however. As their purpose is to adjust and improve ongoing programmes and activities, they need a lean management structure that can allow swift decision-making to ensure progress. The inter-agency RTEs have tended to have one lead agency to take care of day-to-day management, while inter-agency structures have been formed for consultations and discussions. This is a rather new form of evaluative practice, and the model is still being reviewed to identify the best ways of organising inter-agency RTEs.[40]

With RTEs, it is very important to take into account not only how the relationship between the institutions should be, but also the relationship between headquarters level and the field, and between field headquarters and regional/ local branches. A study undertaken by Peta Sandison (2006) has indicated that operational staff are generally more negative to evaluations than are headquarters personnel, because of the extra burden they experience, and that they often do not see evaluations as being a useful tool for their daily work.

A complicating factor is the fact that institutions differ in the distribution of decision-making for policy, planning and operational activities. While some institutions are highly hierarchical, and require that most decisions be endorsed by headquarters, others have decentralised structures where most strategic, programmatic and operational decisions are taken in the field. These variations do not necessarily follow types of institutions, as there are huge differences among UN agencies, among the various NGOs, and also among diplomatic representations to what field headquarters can commit to.

Creating dispute mechanisms
Evaluations should ideally create better-shared understanding of future direction of operations and programmes. In practice, however, most evaluations will inevitably reveal flaws, failures and mistakes by organisations and individuals. Many organisations therefore resist joint evaluations, to avoid being exposed to criticism that might harm funding and support for ongoing operations. The Rwanda evaluation touched upon highly controversial political and policy issues that created tension with important stakeholders. The exposure and criticism of France's diplomatic and political dealings with Rwanda in the years preceding the genocide led the country to withdraw from the evalua-

---

40 The Inter-Agency Standing Committee is (through OCHA) initiated a study in 2009 to gather systematic lessons from the previous IA-RTEs.

tion. There was also resistance within the UN Secretariat (particularly within DPKO) to share documents and information on how the crisis had been handled. The UNHCR also strongly opposed some of the findings of the team evaluating the humanitarian response. The main response of the Steering Comittee of this evaluation was to give priority to the independence of the evaluators over stakeholder sensitivities. At the same time, very high professional standards were demanded from the evaluation teams. The controversial issues were included in the final report, which contributed to give legitimacy to the findings and conclusions.

Conclusions

Can joint evaluations be of use also in multi- national operations? The cases studied have shown that diverse and independent actors can benefit greatly from joint evaluations. Such evaluations have facilitated the development of common policies and standards, and have helped in bringing a range of organisations and institutions into common learning processes. Whether this can work also in complex peacebuilding environments will depend on the ability to elaborate common ground rules, and the inclusiveness of the process. Keywords are respect for the independence and mandates of the various institutions involved. The more complex the relationship between the actors is, the more important a sound *management structure* will be.

It is difficult to envisage a successful joint evaluation between parties that have had no previous relations. Some sort of coordination mechanisms or networks should be there, to underpin a joint evaluation. The existence of formal and informal policy and evaluation networks was seen to be conducive to reach common grounds and facilitated the selection process and the establishment of joint standards in the selected cases. This will be of a greater challenge in situations where also military and political actors are prominent, and where comprehensive coordination structures are lacking or of a patchy nature. Joint evaluations should thus be seen as one of several initiatives to improve coordination.

The credibility of evaluations is closely linked to the independence and transparency of the evaluation process itself. The potential parties would need to reflect on whether they are prepared to be exposed to criticism and scrutiny. This is a relevant point for all actors, but a particular issue for military institutions is the degree to which they can allow independent evaluators the necessary access to information.