

ALNAP Working Paper



Harnessing the Power of
Evaluation in Humanitarian Action:
An initiative to improve understanding and use of evaluation

Alistair Hallam

June 2011

TABLE OF CONTENTS

1. Introduction	3
Background	3
Limitations	4
Outline of this paper	4
A framework to improve understanding and use of EHA	4
Acknowledgements	6
1. Leadership, culture and structure	7
Ensure leadership is supportive of evaluation and monitoring	7
Promote an evaluation culture	7
Increase the internal demand for evaluation information	8
Create organisational structures that promote evaluation	9
2. Evaluation purpose and policy	10
Clarify the purpose of evaluation (accountability, audit, learning)	10
Clearly articulate evaluation policy	11
Ensure evaluation processes are timely and form an integral part of the decision-making cycle	12
Emphasise quality not quantity	12
3. Evaluation processes and systems	14
Develop a strategic approach to selecting what should be evaluated	14
Involve key stakeholders throughout the process	14
Use both internal and external personnel to encourage a culture of evaluation	16
Improve the technical quality of the evaluation process	16
Assign high priority to effective dissemination of findings, including through new media (video, web)	17
Ensure there is a management response to evaluations	18
Carry out periodic meta-evaluations and evaluation syntheses, and review recommendations	19
4. Supporting processes and mechanisms	20
Improve monitoring throughout the programme cycle	20
Provide the necessary human resources and incentive structures	20
Secure adequate financial resources	21
Understand and take advantage of the external environment	22
Next steps	24
References	24

1. INTRODUCTION

Background

ALNAP has done more work on evaluations of humanitarian action (EHAs) during the past few years than at any other time in its history. This has included research on evaluation utilisation, joint evaluations, a pilot guide on real-time evaluations, a meeting and study on impact assessments, a new look Evaluation Reports Database, and the new State of the Humanitarian System report, which included an extensive evaluation synthesis.

However, despite these efforts, and some real improvements within agencies in how they carry out evaluations, there remains a feeling among many in the sector that the full potential benefit of humanitarian evaluations is not being realised. Too often, humanitarian evaluations exist as a disconnected process, rather than becoming embedded as part of the culture and mindset of humanitarian organisations. The sector has made real advances in the quality of evaluations; it would seem that the challenge now is to ensure that evaluations are actually used to improve operational performance.

ALNAP is not alone in coming to such conclusions. The UN Evaluation Group (UNEG)¹ and the wider OECD-DAC group² have been carrying out important work on improving approaches to evaluation for some years across the UN system, and also carrying out peer-reviews of members' evaluation systems and structures. Following a UNEG peer-review, the Office of Evaluation of the World Food Programme (WFP) commissioned work to enhance the learning purpose of the organisation's evaluations (WFP 2009). Similarly, the UK's Department for International Development (DFID) has recently commissioned a study from the Overseas Development Institute on how to improve its learning from both evaluation and research (Jones and Mendizabal 2010). UNICEF has carried out similar work (Stoddard 2005).

There have been similar debates on evaluation effectiveness within the development community for decades. That the humanitarian community is coming to these issues relatively late may reflect the simple fact that it also came later to evaluation itself. However, there is perhaps also something different about humanitarian aid in the scale and inter-connectedness of its ambition, and the subsequently greater difficulties of understanding and attributing impact.

Against this background, ALNAP has commissioned a process of action research, of which this paper is a part, focusing on strengthening institutional understanding as well as capacities and processes in order to better harness the power of EHA. The initial research used literature review and interviews across the ALNAP member organisations and identified areas for further examination. A central emerging lesson is that improving humanitarian evaluation capacities and processes is a complex undertaking that involves a wide range of actors and factors – including allocation of resources;

1. <http://www.uneval.org/>

2. http://www.oecd.org/department/0,3355,en_2649_34435_1_1_1_1_1,00.html

engaging senior leaders; and realigning operational procedures.

This paper starts the process of developing the ideas generated through the literature review and interviews. It is hoped that by illuminating the range of significant factors involved, this paper will help ALNAP members interested in strengthening their evaluation efforts by helping them to identify priority areas of concern, share ideas across the membership about what has and has not worked, and develop new strategies for tackling longstanding issues.

A draft of this paper was presented at a workshop in September 2010 that brought together evaluators from UN, NGO, Red Cross and donor organisations. The paper was subsequently redrafted to take into account comments made at that workshop. The paper was also peer-reviewed by ALNAP members and changes were made to the draft as a result of the helpful and insightful comments received. Publication of this paper represents the end of Phase I of this project.

Limitations

The work discussed in this paper is preliminary, and remains a work in progress. The interviews informing this paper generally involved only one or two people per organisation – usually evaluation specialists. In future, we would need more input from programme, field and management personnel.

Therefore, this paper is a stepping-stone to a broader understanding of trends and issues in strengthening organisational capacities, processes for, and understanding of the evaluation of humanitarian activities. It is intended as a staging post on a route to opening discussion and asking the kinds of questions which will help to advance the debate in pragmatic and practically useful ways.

Outline of this paper

The remainder of this Introduction presents a simple draft framework for increasing the impact and use of evaluations of humanitarian action. This framework takes inspiration from general literature on evaluation utilisation and capacity building, but is modified according to the specifics of the humanitarian context.

The framework is then ‘populated’ in Chapters 1 to 4 with discussion on the specific issues that seem most likely to increase impact and utilisation – as drawn from the literature review and from interviews with a small sample of ALNAP members. The concluding ‘Next steps’ section outlines plans for continuing and developing the work presented in the preceding chapters.

A framework for improving understanding and use of EHA

A draft analytical framework is presented here as an aid to further analysis of what influences understanding and use of EHA. The purpose of the framework is to help shape discussions among

ALNAP members and others involved in evaluating humanitarian action on how to address the problems of poor evaluation impact. The framework draws on literature on capacity development, and was refined following discussions with humanitarian evaluators in order to make it more appropriate for use by those involved in EHA.

The framework is a work in progress. Framework categories that do not add much will be removed, while others will be added as needed. The aim is to keep the framework as light and as easy to use as possible, with a clear focus on the key issues. The aim is also not to be overly prescriptive. At times, where there appears to be consensus from interviews and the literature, the discussions within the framework point towards a solution. However, elsewhere, the answer to a problem may be very agency-specific, when an exploration of the pros and cons of different approaches is more appropriate.

The bare bones of the framework are presented here, with discussion of each capacity area in the four chapters that follow. Overarching issues of leadership, culture and structure are presented first, followed by issues of purpose and policy. Next are more specific evaluation processes and systems. The fourth and final capacity area is supporting processes and mechanisms, some of which are of wider relevance, and do not relate solely to evaluations. Effective capacity building requires attention at all levels of the framework, with much inter-dependence between them.

Capacity Area 1: Leadership, culture and structure

- Ensure leadership is supportive of evaluation and monitoring
- Promote an evaluation culture
- Increase the internal demand for evaluation information
- Create organisational structures that promote evaluation

Capacity Area 2: Evaluation purpose and policy

- Clarify the purpose of evaluation (accountability, audit, learning)
- Clearly articulate evaluation policy
- Ensure evaluation processes are timely and form an integral part of the decision-making cycle
- Emphasise quality not quantity

Capacity Area 3: Evaluation processes and systems

- Develop a strategic approach to selecting what should be evaluated
- Involve key stakeholders throughout the process
- Use both internal and external personnel to encourage a culture of evaluation

- Improve the technical quality of the evaluation process
- Assign high priority to effective dissemination of findings, including through new media (video, web)
- Ensure there is a management response to evaluations
- Carry out periodic meta-evaluations and evaluation syntheses, and review recommendations

Capacity Area 4: Supporting processes and mechanisms

- Improve monitoring throughout the programme cycle
- Provide the necessary human resources and incentive structures
- Secure adequate financial resources
- Understand and take advantage of the external environment
 - Use peer networks to encourage change
 - Engage with media demands for information
 - Engage with donors on their evaluation needs

Acknowledgements

Many thanks to everyone at the ALNAP Secretariat for their support. John Mitchell, Ben Ramalingam and Yuka Hasegawa all gave considerable time to orientating the work and commenting on drafts. Josh Harris helped to set up the pre-workshop questionnaire, and Colin Hadkiss helped with the organisation of the workshop. Thanks to the many ALNAP and non-ALNAP members who agreed to be interviewed. Thanks, in particular, to Rachel Bedouin, Aine Hearn and Karen Proudlock, who gave detailed comments as formal or informal peer-reviewers.

1. LEADERSHIP, CULTURE AND STRUCTURE

Ensure leadership is supportive of evaluation and monitoring

A high proportion of the interviews carried out for this study identified leadership as key in improving the impact of EHA. This is reflected in the academic literature as well as in reviews of the evaluation functions of various humanitarian organisations. Interviewees commented on their experience of a change in leadership having a profound and positive impact on the value and effectiveness of evaluations. Where leaders are not interested in evaluation, or are overly defensive about the performance of their organisations and hence reluctant to accept evaluation findings, a culture develops against learning from experience. If data and analysis are not valued at senior level, this can permeate throughout the organisation and lead to reluctance even to collect the necessary information in the first place.

Even without supportive leadership, an evaluation champion can show the wider organisation the benefits that can arise from evaluation. Good quality evaluations that highlight important issues within programmes that would otherwise have been missed can serve as a tool to convince managers of the importance of evaluation. The board or funders of an organisation can seek to ensure that recruitment of senior managers emphasises the importance of evaluation. A recent study on strengthening learning from research and evaluation in DFID (Jones and Mendizabal 2010) suggested that a director be appointed as 'knowledge and learning' champion.

Promote an evaluation culture

Proudlock (2009) notes that the evaluation literature (eg Patton 2008; Mayne 2008) suggests that culture, as well as leadership, are a key determinant of evaluative capacity. Ideally, an organisational culture will actively seek information on performance in order to learn how to improve the management and delivery of its programmes and services, and to improve its performance. There needs to be a commitment to using evaluation data and then changing behaviour based on that data.

Enabling credible evaluation that is useful and gets used requires establishing effective systems and structures, such as organisational policies, strategic planning, programme reporting and information systems, budgeting, human resources, accountability frameworks and learning processes. Both Patton (2008) and Mayne (2008) point out, however, that this is only ever secondary to establishing the two key elements of culture and leadership. According to Mayne (2008), efforts to create evaluation systems without addressing organisational culture or leadership are likely to end up as burdensome and could potentially work against a culture of seeing evaluation (or any results data) as valuable and worth pursuing.

How do organisations go about cultivating a positive internal culture? Although, as above, evaluation experts state that culture must come first, and then systems, it also seems that developing effective systems and structures for credible evaluations is an important part of the process of changing the culture. Success breeds success, so that high-quality evaluations, that address real information needs, improve the way in which evaluations are regarded. This positively affects the evaluation culture, which, in turn, increases the demand for evaluations, and makes it more likely that evaluation systems and structures are improved in a sustainable way. A virtuous circle ensues. Tackling issues at each stage of the framework identified in this paper can help in creating that virtuous circle.

While many of the building blocks for increasing evaluation capacity will be similar for any organisation, each organisation already has its own unique culture and way of doing things. Therefore, each will need to analyse its own current situation, its readiness for change, and the challenges to building evaluation capacity. Different strategies will then be required for each organisation. Those promoting change will need to be opportunistic, seizing the ‘easy wins’ first, so that impact is quickly seen and support is gathered for the challenge of changing the evaluation culture.

Organisations also stand to learn from the attempts – both successful and unsuccessful – that similar organisations have made to improve evaluation utilisation. Further development of the framework presented in this study will involve workshops, case studies and joint work by small groups of ALNAP members, to help facilitate the process of peer learning.

Increase the internal demand for evaluation information

Increasing demand for information in order to make management decisions will increase the demand for evaluations. But how does one increase demand for information? Patton, writing on utilisation-focused evaluation (2008), talks about the need, when considering potential stakeholders and possible uses, to move from the general and abstract to the real and specific. Very early on in the process, one needs to identify primary intended users and their explicit commitments to concrete, specific uses. In Patton’s view, the evaluator facilitates judgement and decision-making by intended users, rather than acting as a distant, independent judge.

Intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings. This is more likely if users have been actively involved in every step of the process. Patton suggests that evaluation should start with the generation by end users of questions that need to be answered. Such questions should identify information that intended users would like, and which would make a difference to what they do.

Patton refers to research identifying the ‘personal factor’ as one of two key factors that emerged as consistently important in explaining evaluation utilisation.

The personal factor is the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates. Where such a person or group was present, evaluations were used; where the personal factor was absent, there was a correspondingly marked absence of evaluation impact. (Patton 2008)

For Patton, the evaluation process must discover the key stakeholders, namely those who really want to know something. Formal position and authority may matter less than enthusiasm and interest. Evaluators need to nurture and sustain interest, and this requires skills in building relationships, facilitating groups, managing conflicts, and effective interpersonal skills – as well as more ‘technical’ evaluation skills.

These findings are echoed in a recent study on strengthening learning from research and evaluation within DFID evaluations:

the lack of ownership of large evaluations or research programmes delivered externally can help explain their lack of influence. Related to this, direct interpersonal relations between staff and both researchers and evaluators, as well as to the research and evaluation processes, matter a great deal. (Jones and Mendizabal 2010)

Create organisational structures that promote evaluation

There are several ways of organising the management of evaluations. In some organisations, a separate and/or central evaluation unit is responsible for carrying out evaluations; in others, responsibility for evaluation is decentralised throughout the organisation (Hallam 1998). The type of management structure influences, for good and bad, the impact of EHA. For example, having a central unit dedicated to evaluations might improve the quality of evaluations but could undermine broader institutional learning.

There is significant debate about to whom an evaluation department should report. Some organisations have an evaluation department that reports directly to the Board, or to the Chief Executive. Other organisations place the evaluation department lower down in the hierarchy. While there is no right or wrong approach, it is important for organisations to recognise the impact that institutional structure has on learning and accountability. The IMF has an independent evaluation unit answerable to the Board. However, a recent evaluation of the unit found that IMF managers were very resistant to the Evaluation Department (personal communication).

2. EVALUATION PURPOSE AND POLICY

Clarify the purpose of evaluation (accountability, audit, learning)

Whether an evaluation can simultaneously meet both accountability and learning needs has been the most contentious issue arising from the interviews held so far. The literature tends to suggest that these two aims are in conflict, and many interviewees agree. Indeed, a brainstorming session among experienced humanitarian evaluators at a recent training course on evaluation found that an accountability focus in evaluations was a factor in their poor utilisation (personal communication).

Patton notes that an evaluation required by a funder often becomes an end in itself – to be done because it is mandated, not because it will be useful. Mandated evaluations can undercut utility by making the reason for the evaluation compliance with a funding requirement, rather than genuine interest in being more effective (Patton 2008). The recent ODI study on learning lessons from research and evaluations within DFID (Jones and Mendizabal 2010) quotes a similar finding: ‘there is generally a tension between the independence of evaluation departments and their success in engaging users of evaluation’ (Foresti 2007).

Others, however, are equally adamant that any evaluation must address both aims, and that one can only truly learn by being held rigorously accountable. The ‘truth’ will vary by organisation and culture, and one cannot be prescriptive about such things. Nonetheless, what does seem to be lacking in some organisations is clarity about the purpose of evaluations, and recognition that different approaches may be required for different evaluation aims. For example, the peer review of the evaluation function at WFP observes that the ‘tension and complementarities between evaluation for accountability and for learning seem not to be acknowledged everywhere in WFP’ (Baker et al 2007).

Some argue that real learning and performance improvement requires frank and open discussion within an agency, an ability to share failures as well as successes, and a non-judgemental approach. When an evaluation has been externally commissioned, and is carried out by teams of outsiders, such an approach may be particularly difficult to achieve. Some NGOs, that have not been advocates of the independent, accountability-style approach to evaluation, have, however, fully adopted more internal evaluation approaches, which can still involve a considerable degree of criticism. Such criticism is more readily accepted when it comes from insiders who understand the organisation well, and who have themselves been involved in managing similar programmes. Some organisations have found that honest and frank discussion is constrained where evaluations are to be made available to the public.

One way to mitigate the tension between the differing aims of evaluation is to separate 'accountability' evaluations from 'learning' evaluations, and not try to meet all agendas with one exercise. Arguably, it might even be better to call the two types of evaluation by different names, and have the processes managed and run by different individuals and departments within the organisation. Few people enjoy or welcome financial audits, or necessarily expect to learn from them, but no one thinks they should not happen. In a similar way, perhaps, accountability evaluations should be carried out by the audit and accountability department, whereas field-led, learning evaluations should be carried out by a department of knowledge management and learning.

This separation may already be happening to some degree. The UK Disasters Emergency Committee (DEC) now has an Accountability and Audit Manager, as well as someone responsible for learning. A recent DEC report refers to 'separating out and strengthening learning'. At SIDA, there is an aid watchdog as well as an internal evaluation unit, and the internal department has a major focus on utilisation-based learning evaluations. DFID may be heading in a similar direction. Once an explicit separation has been made between accountability and learning evaluations, it then becomes easier to develop a strategy to improve both approaches.

Clearly articulate evaluation policy

EHA policies and guidelines refer to the overall purposes and objectives of the evaluation function for a particular agency and should be where all EHA originates. However, although many organisations have evaluation guidelines, not all distinguish between development evaluation and EHA. Ideally, organisations should have a distinct set of detailed policy guidelines for EHA, as agencies without an agreed policy sometimes experience confusion regarding the evaluation unit's mandate and/or roles (Foresti 2007).

Those promoting change may need to start with developing the appropriate policy, but they will then need to take a strategic approach to rolling this out throughout the organisation. When writing EHA policy guidelines, it is important to consider impact and utilisation from the outset. Arguably, utilisation should be the first aim of any evaluation policy. This will increase the likelihood that the entire evaluation process is designed around end users and their needs.

The organisational challenges of translating policy into practice are not limited to high turnover of staff and poor knowledge management, and are often underestimated by evaluators. For many reasons, policies are not always followed, and it is naive to expect anything different. Where specific EHA guidelines do exist, they should be made available to evaluation teams. Failure to do this can result in sub-standard and even unusable work being produced (Stoddard 2005).

Ensure evaluation processes are timely and form an integral part of the decision-making cycle

A common complaint from potential evaluation users is that evaluations often arrive too late to be of any use in decision-making. A recent report on the way DFID learns from research and evaluations notes:

The most common criticism of evaluations among the interviewees was timing: although they mark important rhythms for Country Offices, they were seen to generally take too long to be relevant to policy teams, and insufficient attention is paid to tying them into policy cycles and windows of opportunity. (Jones and Mendizabal 2010)

This is even more the case for humanitarian evaluations. It is not unusual for evaluations to finish (or even start) only after the peak of the emergency is over. Not only is significant analysis made more difficult by the passage of time, as the memories of key informants wane and information is lost, but long delays between programmes and evaluation conclusions means that there is then no opportunity to use the findings to change the programme concerned.

As a result of concerns such as these, there has been a growth in the number of real-time evaluations (RTEs) by a range of humanitarian organisations. A recent RTE of the Haiti Earthquake (Groupe URD 2010) was able to undertake a field mission within one month of the earthquake, and to capture lessons about the immediate response. It was also able to feed back to actors in the field, either bilaterally or to the cluster groups.

Some evaluation departments work to their own timetables, planning and commissioning evaluations without regard to the management processes that actually influence programmes. Evaluations that appear after major decisions have been made are facing an uphill struggle if they are to have any impact. To have a better chance of bringing about change, evaluation timetabling should start with an analysis of programme planning cycles, and ensure that evaluation products feed into this.

Emphasise quality not quantity

There is finite capacity within any organisation to commission, implement and then learn from evaluation. Some organisations commission evaluations for a large proportion of their programmes, but then find themselves struggling to ensure the quality of the process. Even where quality is maintained, real reflection on the findings of an evaluation report takes time, and considering and implementing recommendations is even more demanding.

ICRC used to request all interested parties to identify the range of questions and issues they would like included in an evaluation. The evaluation department then reframed this into 'evaluable' questions. However, it was found that the scope of the evaluation always grew, until it became difficult to manage the process. To mitigate this, the evaluation department now tries to focus on just three key questions for each evaluation.

Interviewees reported that they feel that too many (often poor-quality) evaluations are being carried out, and that this works against the development of a true evaluative culture within an organisation. Where reports are of insufficient quality, they lose credibility, and the evaluation process becomes less valued in the eyes of managers and implementers. When the same recommendations go unheeded time after time, because of a lack of capacity to debate and act on them, no one takes them seriously. A perception develops that the evaluations are done only for appearance's sake. Sometimes, a quality-focused approach is made more difficult if evaluations are required by donors as a funding condition, forcing organisations to commission more evaluations than they can absorb. Donors can be similarly overwhelmed by the reports that they commission.

3. EVALUATION PROCESSES AND SYSTEMS

Develop a strategic approach to selecting what should be evaluated

Some evaluation departments review their organisation's annual expenditure and plan evaluations on the basis of ensuring that all major programmes are evaluated every few years. The guiding principle of such approaches is that accountability demands coverage of the organisation's activities on a cyclical basis. However, such a mechanistic approach does not necessarily lend itself to ensuring effective utilisation and impact of EHA.

There may be little that is controversial about large programmes that have been running for years, and evaluations are unlikely to lead to any radical changes. There may also be all sorts of barriers to change in some areas, where evaluation findings are likely to be ignored. For accountability reasons, evaluations are often required of large-budget programmes. A methodology that is lighter on the learning aspect might be more appropriate when such programmes are uncontroversial. (This underlines the importance of the need to clarify the purpose of the evaluation at the outset, as noted in Section 2 above.)

A strategic approach to selecting what should be evaluated is fundamentally different from the 'cyclic' approach described above, and starts by looking at how the evaluation process can add value. As a good example, SIDA's evaluation planning cycle starts with the evaluation unit having conversations with all operational units, to determine their knowledge needs, what they would like to know and how evaluation could help. A list of around 100 ideas for evaluations is initially generated, from which the evaluation department chooses the 15 they will carry out.

Evaluations not chosen are subject to decentralised evaluations, on which the evaluation department gives feedback and advice. Staff members who proposed any of the 15 selected evaluations form reference groups for that evaluation. They must list the intended use and users. If there are not enough users, the evaluation is dropped. Otherwise, these intended users are then involved in drafting terms of reference, and work with the reference group throughout (from a presentation by Joakim Molander, SIDA, at DFID workshop on lesson learning, 22 June 2010).

Involve key stakeholders throughout the process

Evaluations are political and concern various interest groups (Carlsson et al, 1994, in Sandison, 2006) and stakeholders, both internal and external. The most important external stakeholders are the affected populations themselves, who should arguably be at the centre of not just the evaluation

process, but the whole humanitarian system. Although there is still a long way to go for this to be achieved, a large number of humanitarian organisations are working to improve the current situation. It is now slowly becoming the norm for 'beneficiaries' to be involved at some stage of the evaluation process, although this is usually only as informants rather than in setting the evaluation agenda. Nonetheless, this represents an improvement on the situation of a decade ago, when few humanitarian evaluations involved structured discussions with the affected population at all.

Despite recent advances in 'downward' accountability, the Humanitarian Accountability Programme's recent annual report³ finds that considerably more needs to be done with respect to evaluation practice. However, given that the evolution of EHA to date has been driven largely by greater demand for 'upward' accountability to donors, efforts to reorient the evaluation process towards the concerns of disaster-affected stakeholders may require entirely new mechanisms: it may not be possible simply to use existing mechanisms to fulfil new and different functions.

This paper does not seek to duplicate discussions on humanitarian accountability being conducted elsewhere. However, as a minimum, all evaluations should ensure that beneficiaries are enabled to contribute, with their ideas feeding into analysis, learning and recommendations. A large literature exists on a range of methods for obtaining a valid spectrum of beneficiary views, which is not to say that this is easy to achieve. There remain very considerable challenges in ensuring true participation in situations where power structures prevent evaluators from accessing certain voices, typically those of women and minority groups. This is exacerbated in conflict situations where expressing opinions might be not only difficult but also dangerous.

Other external evaluation stakeholders include the general public, the media, various governance structures, and donors. FAO systematically involves stakeholders from the beginning of thematic or country evaluations, using consultative groups including affected people, major donors and programme personnel. The stakeholders form an advisory group, and build trust and 'buy-in' – considered essential by the FAO Evaluation Unit in achieving change through evaluation. The group ensures that the right questions are asked, and has a chance to comment on draft products.

Research on strengthening learning from research and evaluation in DFID (Jones and Mendizabal, 2010) found that increasing the sense of ownership of evaluations by DFID staff members was essential. Where interviewees had been involved actively with the DFID Evaluation Department in one way or another, this improved their understanding of the department, breaking down any misconceptions about being 'audited', and was felt to be valuable. On the other hand, concerns were noted that lessons from evaluations relevant to DFID more generally, at a strategic level, had no 'owner', and were thus less likely to be acted upon. UNHCR has a steering committee for each evaluation that includes staff members from the relevant units.

3. <http://www.hapinternational.org/pool/files/a-the-2009-humanitarian-accountability-report.pdf>

Use both internal and external personnel to encourage a culture of evaluation

A review of the literature and discussions with field staff reveals a common theme of field and operational staff being unhappy with current evaluation practices. Many field personnel feel that evaluations take up valuable time, and that they end up 'teaching' the evaluators about the programme and the issues, and then subsequently learn little from the reports, if they ever see and read them. Furthermore, they frequently find that nothing seems to change as a result of the evaluation. There is also concern that outsiders need a lot of time to learn about the culture and practice of the organisation being evaluated. Recommendations can be unrealistic where this learning has not taken place, damaging the credibility of the evaluation.

A number of interviewees mentioned that they have started using more internal evaluation, or mixed teams of insiders and outsiders. The involvement of insiders means that findings and recommendations are more likely to be appropriate. In addition, insiders are more likely to have a better understanding of the concerns of field personnel, and of their perspective on key issues. Using internal evaluators also brings the enormous benefit of retaining the experience and knowledge gained by those carrying out the evaluation. It is expensive to hire external evaluators to learn huge amounts about the organisation and the emergency response, only for them to take away this knowledge when their work is done.

Despite the many benefits of using insiders, some organisations are implacably opposed to doing so, fearing that this will limit independence and restrict any radical recommendations that might be required on restructuring the organisation's response. As noted above, it is essential to clarify the reasons for carrying out an evaluation, and to recognise that trade-offs are inevitable.

Improve the technical quality of the evaluation process

The rationale behind this study on improving evaluation utilisation is that there are many factors affecting the impact of EHA, and that technical quality is only one of these. However, technical quality is clearly very important. High-quality evaluations increase the credibility of the whole evaluation process, and create the potential for a virtuous circle to develop: if evaluations are valued more highly, this creates the right conditions for more of them to be of higher quality in future. ALNAP is commissioning a new Guide to the Evaluation of Humanitarian Action as a separate exercise, and so this important issue will not be discussed further here.

Assign high priority to effective dissemination of findings, including through new media (video, web)

Evaluations are of potential interest to agency managers, beneficiary representatives, local governments, donor officials, students of humanitarian issues, and the press, as well as the general public. Ideally, each audience should receive a different product tailored to its needs (Hallam 1998). However, increasing impact is also about constructing pathways for the evaluation findings to make a difference within the organisation:

performing a good quality evaluation is only the first step. The lessons then have to be absorbed, taken forward, and implemented in practice before organisational learning can be said to have occurred. (Stoddard 2005)

It is often assumed that evaluations yield information or provide lessons learned that flow directly back into the policy cycle and are thus incorporated in the planning of future programmes and projects. In this way, there is a constant learning process leading to ever-improving performance (Frerks and Hilhorst 2002). This understanding presupposes a rational, scientific planning model – which has never been adopted in daily development practice. In a plural, complex and disorderly society, decisions on goals and programmes are often political compromises that do not necessarily correspond with the outcomes of evaluation (Frerks and Hilhorst 2002).

Strategic dissemination of findings is key to making evaluations more effective. In a personal communication, a senior evaluator commented on the difference in the impact of evaluations he had noted when moving from a large and well-resourced evaluation department to a much smaller one in a different organisation. Despite the differences in resources available, the evaluations carried out by the smaller organisation had significantly more impact. There were several reasons for this, but the most important reason was the smaller organisation's commitment to disseminating the results and targeting that dissemination effectively. Indeed, the evaluation process began with the communication strategy, rather than dissemination being thought about only once there was a (perhaps inappropriate) report in hand. Planning ahead influences the type of information collected throughout the evaluation, and ensures it meets the needs of decision-makers.

A recent ODI paper on learning lessons from research and evaluation (Jones and Mendizabal 2010) echoed the above findings, and found that how and when evidence is produced, presented and communicated matters. Options include targeted seminars and presentations, one-to-one briefings for team leaders, an evaluation department newsletter or briefing papers, short email products, and the development of new products such as documents that present lessons from evaluations along thematic, regional/national and programmatic/policy lines. Focusing on personal interrelations rather than intermediaries was also considered important. The same study found that the format and

presentation of evaluations could be improved: 'Some interviewees felt that the full reports were too long and technical... it is likely that this is a common problem with the evaluation profession' (Jones and Mendizabal 2010).

Some agencies have stopped putting recommendations into evaluation reports, because they feel that focus on recommendations can detract from analysis and learning. The most important part of the evaluation for them is the analysis, because this is what encourages learning. It is better that those who know and understand the organisation develop and take ownership of the recommendations. It is also important that evaluation reports and relevant findings are easily accessible. Ways of enabling this include: good cataloguing; improving intranets to make it easier to search and browse, and making them RSS compatible to allow staff to link to them without having to access the site every time they want information; and producing global lessons learned from meta-analyses.

Interviews conducted for this study revealed some innovative dissemination strategies. A number of organisations, including UNHCR, ICRC and Groupe URD, have made documentaries around evaluations, some reporting powerfully on the key issues found in the evaluation solely in the words of beneficiaries. ALNAP itself produced a Lessons Paper on humanitarian activities after earthquakes, by reading and distilling the lessons from a number of evaluations of earthquake response (Cosgrave 2008). This paper was downloaded over 3,500 times within days of the January 2010 earthquake in Haiti.

Ensure there is a management response to evaluations

Management response and follow-up to evaluations is a key area for improving the impact of evaluations. There are numerous anecdotal reports of evaluation reports gathering dust, largely unread, and with little formal or informal follow-up on findings and recommendations. This may be partly because of problems noted above – particularly lack of clarity of purpose of the evaluation, and lack of involvement by key stakeholders. Nonetheless, formal systems of management response to evaluations serve to reduce the chance that an evaluation process ends with the production of the report.

Different organisations take different approaches to this issue. UNDP has created an Evaluation Resource Centre. This is a public platform, designed to make UNDP more accountable, where evaluation plans are logged along with a management response. Responses and follow-up are tracked by the evaluation department and reported to the Executive Board of UNDP. Some DFID country offices hold an 'in day' to go over performance frameworks and evaluation results, to make sure that key lessons are not lost.

In FAO, there is a process in which senior management comments on the quality of evaluation reports, as well as on what findings they accept and what actions are planned to address these.

This feedback is presented to the governing body along with the evaluation. In addition, for major evaluations, there is a further step in the process: two years after the evaluation, managers are required to report to the governing body on action taken on the recommendations they accepted at the time. For FAO, these have proven to be powerful management tools, and which include the opportunity to revisit evaluation findings, and to have a dialogue about the management response to them.

Carry out periodic meta-evaluations and evaluation syntheses, and review recommendations

The huge demand following the 2010 Haiti earthquake for ALNAP's Lessons Paper Responding to Earthquakes: Learning from earthquakes relief and recovery operations (Cosgrave 2008) has been mentioned above. It seems there is a significant demand within the humanitarian sector for such meta-analysis and thematic reviews.

- Oxfam has recently started to use the same six benchmarks for its internal reports. Every few years, a meta-evaluation is carried out against these benchmarks 'which is incredibly valuable' (personal communication).
- CARE has carried out a meta-review of evaluations and after-action reviews from 24 emergency responses, and drawn together key learning from this.
- Every year, NORAD produces a synthesis report of lessons from evaluations.
- The ODI paper on learning lessons from research and evaluations (Jones and Mendizabal 2010) found glowing references from interviewees to research and evaluation that offered synthesis and comparison of work from around DFID on particular themes or sectoral areas.

Such meta-approaches and syntheses are important in extracting full value from expensive evaluation processes. They help to ensure that findings across many different evaluations are validated and are not specific to just one project. Greater consistency of findings across programmes leads to more confidence in their credibility, and so to greater potential impact of making changes on the basis of such findings.

4. SUPPORTING PROCESSES AND MECHANISMS

Improve monitoring throughout the programme cycle

Evaluators frequently find that a lack of monitoring data significantly impedes their ability to make meaningful comments on the effectiveness of a humanitarian programme. At times, managers seem to request evaluations to compensate for the lack of monitoring information. This is an expensive way of obtaining information that should be routinely collected and analysed.

The lack of monitoring data is a long-standing problem, due partly to the nature and circumstances of humanitarian work: a multiplicity of actors, the fluidity of the situation, the difficulties of working in the context of war and instability, and the frequent absence of baseline data on the previous condition of the affected/target populations. However, there are also attitudinal constraints, where emergency personnel do not consider it a priority to set up monitoring systems from the outset. There is also a lack of agreement on standardised monitoring procedures and protocols among agencies, despite widespread adoption by many humanitarian agencies of Sphere standards. Indeed, several interviewees suggested that all humanitarian programmes should report against Sphere standards.

Standardisation of data collecting and reporting would make it easier for the system as a whole to see where gaps exist in humanitarian provision. If agencies and donors could agree on what data should be collected, this would significantly increase the likelihood of it happening, meaning that evaluators could be more confident that they will find the data they need to carry out their work. A disadvantage of such an approach would be that there might be a tendency to focus on the measurable and pay less attention to contextual analysis, already a weak point in some humanitarian evaluations. Context-specific monitoring would always be required as well, to ensure that a programme was meeting its stated objectives, although this would not prevent standardisation in many areas.

Provide the necessary human resources and incentive structures

‘The incentive structures of... agencies do not necessarily reward those in the evaluation departments, which, as a result, are not able to offer clear career opportunities for staff members’ (Foresti 2007). A WFP report made similar comments about career incentives for staff in evaluation departments: [evaluation office] posts are highly stressful regarding relations between staff and other members of the organisation ... Some felt that a posting in [the evaluation office] was a ‘bad career move’ and might affect the individual’s future career. (Baker et al 2007)

Given such incentive structures or lack of them, it may be that evaluation departments will not perform at optimal levels.

As well as formal incentive structures, there are also many informal incentives that can either promote good evaluation or act as an obstacle to it. Such incentives can be subtle to outsiders, but play an important role in determining whether evaluations are likely to be used effectively. For example, where managers demand an evidence base before implementing new policies and programmes, evaluations are likely to be more highly valued.

A recent study of lesson learning from DFID research and evaluations (Jones and Mendizabal 2010) found that the demand for an evidence base increased in circumstances where a rapid response was required to an unexpected event. Under these circumstances, the ability of personnel to use evidence and lessons from research and evaluations within their work undoubtedly built their credibility and influence within DFID. In particular, evidence that proved the value for money of a particular policy area was highly prized, as this was useful to DFID leaders in strengthening their case for increased budgets with the UK Treasury (underlining the importance of messages 'from the top').

However, in general, the study found that original thought and new ideas were more likely to be rewarded than the use of proven ideas or lessons learned elsewhere. Orienting incentives towards 'new thinking', whilst understandable, is likely to militate against the effective use of evaluation. The same study also found that time pressures encouraged staff members to rely on experience, rather than evidence based on research and evaluation, to support their arguments.

The informal incentive structures within an organisation are part of its culture, and merely changing formal structures might not tackle the underlying and less easily measured constraints to effective evaluation. Strong leadership, which supports the use of evidence and evaluation, is important in bringing about change in such circumstances. However, it is also important not to neglect the formal incentive structures and processes. Evaluation staff should be trained to an appropriate degree, and have access to continuing education on evaluation. Staff rotation through the evaluation department should be encouraged and regular.

New members of staff should learn about the evaluation system when they are inducted into the organisation. As discussed above, using a mix of insiders and outsiders in evaluation teams can help staff members to learn about evaluation, improve their analytical skills and promote their acceptance of evaluation. Programme planning processes should be designed so that there is time and space to consider and debate evaluation findings, and for recommendations (where accepted) to be adopted.

Secure adequate financial resources

Although many evaluations are driven by issues of cost-effectiveness, the costs and benefits of

evaluation itself are rarely analysed. Research in this area could lead to a better understanding of the concrete benefits of evaluations, proving that 'Doing humanitarian evaluations better through an ongoing process of reflection and learning is the most cost-effective way for ensuring that humanitarian programmes improve' (Apthorpe et al 2001), and thus improving the status of evaluation staff and departments.

There is a widespread perception that joint evaluations are prohibitively costly, yet, provided they are well-designed and managed, their benefits may outweigh these costs. Many field workers and managers complain about the costs of evaluations, yet do not always stop to consider the potentially huge financial and human costs of continuing with unsuccessful programmes (perhaps because evaluations only rarely succeed in influencing such programmes).

In the past, some organisations allocated a fixed percentage of programme costs to evaluation budgets. While this was initially useful as a way of generating resources for evaluations in organisations not used to doing them at all, it is unlikely to be the most cost-effective way of using scarce resources. Better to take a strategic approach, with proportionately more resources going to complex programmes, or programmes that are more contentious and where evidence for effectiveness is less readily available.

Understand and take advantage of the external environment

Use peer networks to encourage change

Peer networks are an essential component of bringing about change. This paper was commissioned by ALNAP, a key network on humanitarian evaluation, in order to promote peer learning around evaluation impact. Peers, by definition, work in similar contexts, and so their experiences can be used by similar organisations to increase their own learning. ALNAP promotes peer learning through workshops and publications, and visits to member organisations.

The evaluation departments of the UN organisations also have a peer grouping known as the UN Evaluation Group (UNEG). UNEG has its own peer review process, which involves visiting one of the organisations and assessing its evaluation function against a set of norms and standards developed by the group as a whole. Members have found the peer review process useful for raising the profile of evaluation within their organisations, and helping evaluations become more embedded within the organisations' processes.

Some organisations have sought to develop their own peer groups. For example, FAO has now institutionalised peer reviews, their Charter of Evaluations demanding and budgeting for a biannual peer review of the organisation's evaluation function.

Engage with media demands for information

The media have a tendency to categorise humanitarian actors as either saints or sinners. Today's major humanitarian responses are characterised by huge media interest, with journalists scrambling to be the first on the scene and then complaining about the slow speed of the response, ignoring the many constraints and challenges involved. This can lead to a defensive response by agencies and unwillingness to acknowledge any problems for fear of criticism. This works against a culture of learning and improvement.

The UK DEC changed its policy on commissioning and publishing evaluation studies following negative media reports after the response to the Southern Africa drought in 2002/03. Ironically, the evaluation report was generally very positive, with relatively few criticisms. Unfortunately, it was these criticisms that were seized on by the press. An arguably more effective response over the longer term would be a greater openness and a willingness to accept that problems occur, and to debate them. This, after all, happens with many other public interventions (health service delivery, for example) without the public losing support for the intervention.

Interestingly, the response to the 2004 tsunami appeal – the one immediately after the appeal for the Southern Africa drought discussed above – was strongly supported by the UK public, with no evidence that any donations were held back because of earlier media criticisms of the same agencies. Agencies should also engage directly with the media to educate them about humanitarian response, to avoid the unbalanced and unhelpful journalism that can accompany a large-scale emergency.

Engage with donors on their evaluation needs

According to Frerks and Hilhorst (2002), EHAs are more often than not donor initiatives driven by donor values and interests. It is, therefore, vital that donors support efforts to improve the impact of EHA. A significant barrier within organisations to developing an evaluation 'culture' is that evaluations are often seen as compliance instruments (as discussed in Section 2 above). Organisations committed to evaluations only as a condition for accessing funding are unlikely to value or have 'ownership' of the resulting products. Consequently, agencies may undervalue evaluations, which can represent poor value for money in terms of the actual learning and performance change that they bring about.

A report on DFID's lesson learning from research and evaluations (Jones and Mendizabal 2010) found that, in countries that were higher HQ and UK political priorities, the profile of evaluation was enhanced. This meant that recommendations were more likely to be accepted (again demonstrating the powerful influence of clear leadership demands on practice). If donors and agencies can agree a strategic approach, then both could benefit from carrying out fewer, more targeted evaluations that focus on key issues rather than a predetermined list of projects and programmes.

Next steps

The purpose of this paper is to stimulate debate on what is required to increase the impact of humanitarian evaluations. From this initial analysis, and subsequent workshops with practitioners, we hope to develop plans to build on this work and bring about meaningful change.

Without doubt, the way forward will include building stronger links between evaluation and programme/management staff. It is undeniable that 'the majority of the characteristics and impediments to organisational change are beyond the control of evaluation' (Sandison 2006). Evaluation must be placed and judged within a larger management framework, for 'if we continue to expect evaluation to cover most of the accountability and learning needs of the sector, we will be disappointed' (Sandison 2006).

The next task will be to develop a self-assessment questionnaire, so that organisations can see how they are doing in terms of addressing the issues that influence evaluation utilisation. This paper, as well as the questionnaire, will provide a framework for generating more detailed case studies of how organisations have improved the utilisation of humanitarian evaluations. Chief executives of ALNAP member organisations will be interviewed, to find out what information senior decision-makers want, where they get this information from, and how evaluations could better serve their needs.

It is vital that we check evaluators' priorities for change against those of operational managers. Without dialogue between the two groups, the potential impact of evaluation as a management tool will not be realised. The case studies and the information generated from further in-depth interviews should significantly enrich the ideas discussed in this paper, and will lead to a revised version later in 2011.

REFERENCES

Cosgrave, J. (2008) Responding to Earthquakes: Learning from earthquake relief and recovery operations. ALNAP Lessons Paper. London: ALNAP.

Apthorpe, R., Borton J. & Woods (eds.) (2001) Evaluating International Humanitarian Action: Reflections from Practitioners. London: Zed Press.

Baker, J. et al (2007) Peer Review: Evaluation Function at the World Food Programme. SIDA.

Carlsson, J. et al (1994) The Political Economy of Evaluation. London: Macmillan Press.

Groupe URD (2010) Real-time Evaluation of the Response to the Haiti Earthquake of 12 January 2010. (available at www.urd.org).

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

- Foresti, M. (2007)** A Comparative Study of Evaluation Policies and Practices in Development Agencies. London: ODI.
- Frerks, G. & Hilhorst, D. (2002)** 'Evaluation of Humanitarian Assistance in Emergency Situations', UNHCR/EPAU: New Issues in Refugee Research, Working Paper 56.
- Jones, H. & Mendizabal, E. (2010)** Strengthening learning from research and evaluation: going with the grain. London: ODI.
- Mayne, J. (2008)** Building an Evaluative Culture for Effective Evaluation and Results Management. Institutional Learning and Change Initiative (ILAC) Brief20. (available at www.cgjar-ilac.org).
- Patton, M. (2008)** Utilization-Focused Evaluation. London: SAGE Publications.
- Proudlock, K. (2009)** Strengthening Evaluation Capacity in the Humanitarian Sector: A Summary of Key Issues. (author's unpublished notes).
- Sandison, P. (2006)** 'The Utilisation of Evaluations' in ALNAP Review of Humanitarian Action: Evaluation Utilisation. London: ODI.
- Stoddard, A. (2005)** Review of UNICEF's Evaluation of Humanitarian Action. New York: United Nations Children's Fund.
- WFP (2009)** Closing the Learning Loop – Harvesting Lessons from Evaluations: Report of Phase I. Rome: World Food Programme.

ALNAP Working Paper



ALNAP

Overseas Development Institute
111 Westminster Bridge Road
London SE1 7JD, UK

Tel: + 44 (0)20 7922 0300
Fax: + 44 (0)20 7922 0399
Email: alnap@alnap.org
www.alnap.org