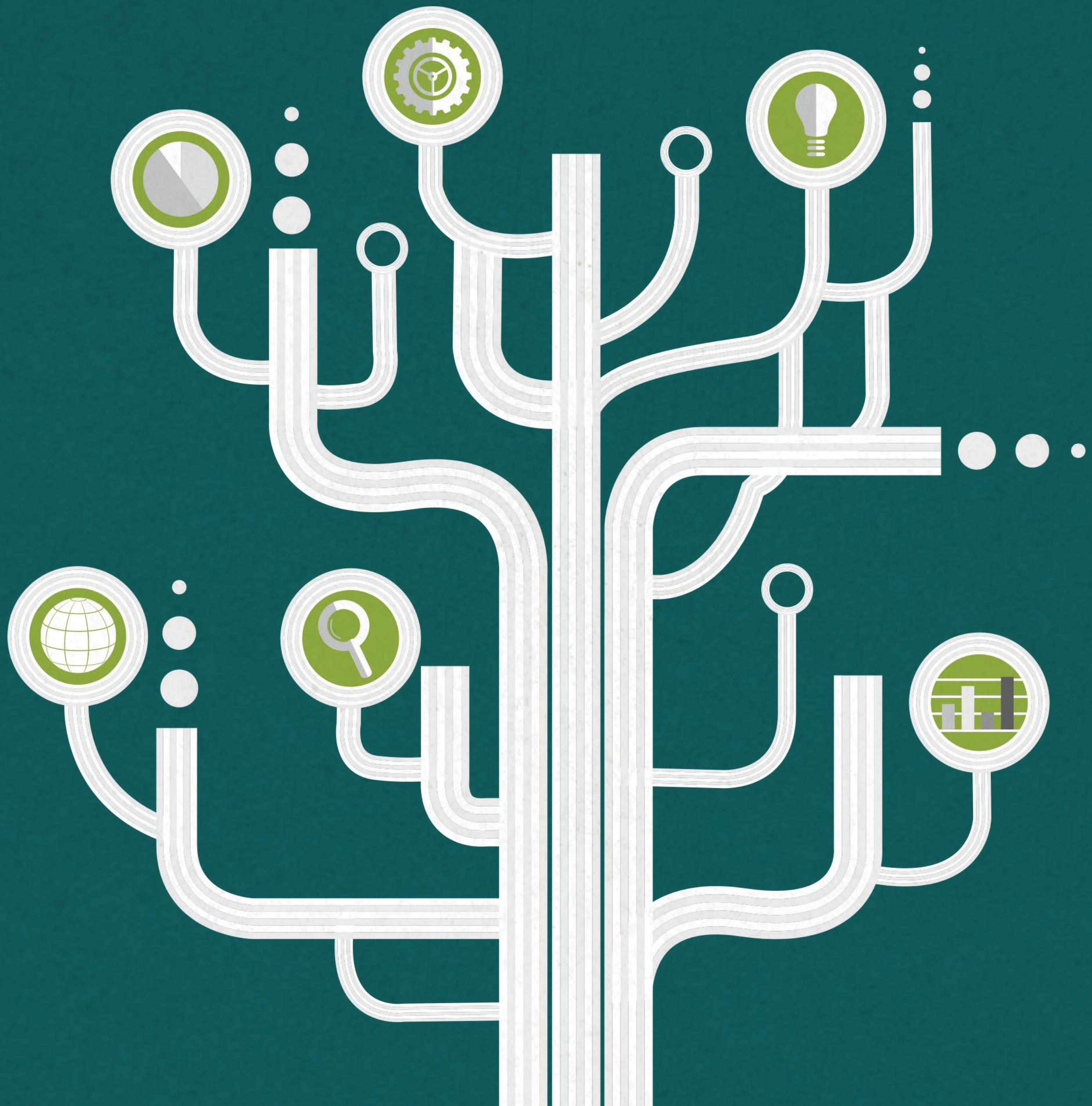
ALNAPSTUDY

USING EVALUATION FOR A CHANGE:

Insights from humanitarian practitioners

Alistair Hallam and
Francesca Bonino



ENTER

ALNAP is a learning network that supports the humanitarian sector to improve humanitarian performance through learning, peer-to-peer sharing and research. A unique network, our members are actors from across the humanitarian sector: donors, NGOs, the Red Cross/Crescent, the UN, independents and academics.

www.alnap.org/using-evaluation

Acknowledgements

Franziska Orphal and Alexandra Warner from the ALNAP Secretariat have been providing invaluable communication and research assistance support throughout this initiative. Paul Knox-Clarke and Patricia Curmi helped with reviewing earlier drafts of this study and advising on content design work, respectively.

Thanks finally to all of those who provided insights, case study material and, most precious of all, their time to debate these issues with us. The list of contributors at the end of this report includes interviewees and those who contributed evaluator insights through the ALNAP Evaluation Community of Practice. Any misrepresentation or shortcomings in the text are those of the authors alone.

We would also like to acknowledge all those who helped with the earlier working paper on this subject, in particular, John Mitchell, Ben Ramalingam, Yuka Hasegawa, Rachel Bedouin, Aine Hearns and Karen Proudlock as well as the many ALNAP and non-ALNAP members who agreed to be interviewed during the first phase of this initiative.

Suggested citation

Hallam, A. and Bonino, F. (2013) *Using Evaluation for a Change: Insights from humanitarian practitioners.* ALNAP Study. London: ALNAP/ODI.

Designed by Deborah Durojaiye
Edited by Nina Behrman

This publication is also available as an e-book and in print. If you have comments or suggestions please email us at eha@alnap.org

ISBN 978-0-9926446-0-4
ISBN 978-0-9926446-1-1 (e-book)
ISBN 978-0-9926446-2-8 (PDF)

About the authors

Alistair Hallam is an economist and a medical doctor with over 20 years of experience working in humanitarianism. He is director and co-founder of Valid International. He has written widely on humanitarian issues, authoring amongst other works, the influential guide to evaluating humanitarian aid programmes, 'Evaluating Humanitarian Assistance Programmes in Complex Emergencies'

Francesca Bonino is Research Officer at ALNAP where she covers the evaluation, learning and accountability portfolio. She earned a Ph.D. having undertaken research on the development of the evaluation function in the humanitarian sector and within humanitarian organisations.

HOW TO USE THIS INTERACTIVE PDF

At the bottom of each page in this PDF you will see a series of icons.

Jump back to previous view	Move to previous page	Move to home page	Move to next page	Move to 'How to use' page	Feedback button

Use the buttons below to jump to the different Capacity Areas.

1 2 3

A complete annotated bibliography is also available on the ALNAP website here

This Interactive guide is best viewed with Adobe Reader. You can download the latest version free at http://get.adobe.com/uk/reader/

This guide can be viewed on tablets and smartphones, but some navigation features may be disabled.

Working offline – No internet connection is required to use the guide: we encourage you to download this interactive PDF and use in Adobe Reader. Feedback and viewing references do require an internet connection.

Capacity Area 1: Leadership, culture, structure and resources

- Ensure leadership is supportive of evaluation and monitoring.
- Promote an evaluation culture:
  - decrease perception of evaluation as criticism and evaluators as critics
  - use both internal and external personnel to carry out evaluations
  - re-brand evaluations
  - be flexible and have fun
  - get incentives right.
- Create organisational structures that promote evaluation.
- Secure adequate resources – financial and human.

Capacity Area 2: Purpose, demand and strategy

- Clarify the purpose of evaluation (accountability, audit, learning) and articulate it in evaluation policies.
- Increase the demand for evaluation information:
  - strive for stakeholder involvement
  - ensure evaluation processes are timely and integral to the decision-making cycle.
- Develop a strategic approach to selecting what should be evaluated.

Capacity Area 3: Evaluation processes and systems

- Strengthen pre-evaluation processes.
- Improve the quality of evaluation:
  - limit the focus of evaluation
  - involve beneficiaries
  - quality assurance
  - engage in peer-review of the evaluation function.
- Disseminate findings effectively.
- Strengthen follow-up and post-evaluation processes including linking evaluation to wider knowledge management:
  - ensure there is a management response to evaluations
  - ensure access to and searchability of evaluative information.
- Conduct meta-evaluations, evaluation syntheses, meta-analysis and reviews of recommendations.

Contents

Boxes

Evaluator insights

Acronyms and abbreviations

3ie	International Initiative for Impact Evaluation
ACF	Action Against Hunger
ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
CIDA	Canadian International Development Agency
CoP	community of practice
CO	country office
DAC	Development Assistance Committee (OECD)
DARA	Development Assistance Research Associates
DEC	Disasters Emergency Committee
DECAF	DEC Accountability Framework
DFID	UK Department for International Development
ECB	Emergency Capacity Building Project
ECB-AIM	ECB Accountability and Impact Measurement
EHA	evaluation of humanitarian action
EQAS	Evaluation Quality Assurance System (WFP)
FAO	Food and Agriculture Organization of the UN
GEROS	Global Evaluation Reports Oversight System (UNICEF)
Groupe URD	Groupe Urgence, Réhabilitation, Développement
HAP	Humanitarian Accountability Partnership
IA RTE	IASC Inter-Agency Real Time Evaluation
IASC	Inter-Agency Standing Committee
ICAI	Independent Commission for Aid Impact
IEG	Independent Evaluation Group (World Bank)
IFRC	International Federation of Red Cross and Red Crescent Societies
INGO	international NGO
M&E	monitoring and evaluation
MSF	Médecins sans Frontières
NGO	non-governmental organisation
OCHA	UN Office for the Coordination of Humanitarian Affairs
ODI	Overseas Development Institute
OECD	Organisation for Economic Co-operation and Development
RC/RC	Red Cross and Red Crescent movement
RTE	real-time evaluation
Sida	Swedish International Development Cooperation Agency
SOHS	The State of the Humanitarian System (ALNAP report)
ToR	terms of reference
UN	United Nations
UNDP	UN Development Programme

UNEG	UN Evaluation Group
UNHCR	UN High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNRWA	UN Relief and Works Agency for Palestine Refugees in the Near East
USAID	United States Agency for International Development
WB	World Bank
WFP	World Food Programme

INTRODUCTION

Background

Evaluation of humanitarian action (EHA) is an area of practice at the core of ALNAP's work. In the last few years, ALNAP has focused its 'evaluative' attention on the development of guidance (Cosgrave et al., 2009; Beck, 2006; Proudlock and Ramalingam, 2009; Beck and Buchanan-Smith, 2008; Cosgrave and Buchanan-Smith, 2013) as well as on issues related to broader, system-wide performance monitoring (see for instance Beck, 2002 and 2004a; Wiles, 2005; Beck and Buchanan-Smith, 2008; ALNAP, 2012). ALNAP members have always shown real interest in discussing and engaging in issues of topical interest such as joint evaluation, impact assessment and real-time evaluations in humanitarian settings. Moreover, Peta Sandison's watershed study on evaluation utilisation (Sandison, 2006) opened an important new stream of joint research and engagement with ALNAP members on the broader theme of evaluation utilisation (Hallam, 2011).

Sandison's study drew on some of the classic literature on evaluation use (Chelimsky, 1997; Pawson and Tilley, 1997; Weiss, 1998; Guba and Lincoln, 1989; Feinstein, 2002; Patton, 1997). For the first time, it brought to the attention of the broader humanitarian sector the issues, opportunities and challenges in getting evaluation used to improve humanitarian practice. Building on that earlier work, the present study aims at engaging ALNAP members to tease out and make explicit some of the most recent thinking and practical attempts to increase the chance that evaluation will be used more effectively to contribute to learning and change in humanitarian practice.

The importance of improving evaluation design, approach and methodology has been extensively discussed in the literature (House, 1983; Scriven, 1991; Grasso, 2003; Kellogg Foundation, 2007; Rogers, 2009; Stern et al., 2012; WB-IEG, 2009). However, ALNAP members felt that producing high-quality evaluation products is only half of the challenge confronting humanitarian actors. The other half relates to strengthening the capacities of individuals, teams and organisations to plan, commission, conduct, communicate and follow-up credible and timely evaluations.

Over the past five to eight years, some real improvements have taken place in the practice of humanitarian evaluation, in the quality of evaluations produced, and in the experience of new ways of integrating evaluation within larger organisational processes (WFP, 2009; Jones et al., 2009; ACF, 2011b; UNICEF, 2013; Tearfund, 2013). Nonetheless, many in the sector continue to feel that the full potential benefit of humanitarian evaluations is not being realised, and that they can be better embedded within the culture, processes and structure of humanitarian organisations. ALNAP research confirms this: opportunities to maximise benefit from evaluations are not always taken and there are barriers to utilising evaluation findings, which need to be overcome (Sandison, 2006).

A recent study on the state of evaluation practice among US charities and non-profits concluded that more than two-thirds of organisations surveyed 'do not have the promising capacities and behaviours in place to meaningfully engage in evaluation'. The same study found that evaluation is an often-undervalued, overlooked tool for improving outcomes and maximising impact: 'nice-to-have, not a need-to-have' (Innovation Network, 2013: 1). Despite some recent progress at single- and inter-agency levels, the above observation may well resonate with some of the reflections often made on evaluation practices in the broader humanitarian sector (see for instance Lindahl, 1998; Crisp, 2000; Griekspoor and Sondorp, 2001; Wood et al., 2001; Frerks and Hilhorst, 2002; Darcy, 2005; Fearon, 2005; Ridde and Sahibullah, 2005; Mebrahtu et al., 2007; Guerrero et al., 2013).

Encouragingly, there are efforts to tackle these issues on a number of fronts. For example, the UK Department for International Development (DFID) has issued a new evaluation policy emphasising the role of evaluation at intervention design stage (DFID, 2013: 2, Paragraph 10). DFID has also been commissioning studies looking at evaluation methods (Stern et al., 2012; Vogel, 2012; Perrin, 2013) and exploring how to improve learning from both evaluation and research (Jones et al., 2009; Jones and Mendizabal, 2010). UNICEF, following a review of its humanitarian evaluation function (Stoddard, 2005), has been working to strengthen the internal quality of its evaluation products (UNICEF, 2010a), while the World Food Programme (WFP) Evaluation Office, following a United Nations Evaluation Group (UNEG) peer-review, commissioned some work to enhance the learning purpose of the organisation's evaluations (WFP, 2009).

The past five years have also seen a notable increase in inter-agency real-time evaluations (IA RTEs), following major crises such as cyclone Nargis (IASC, 2008), the Pakistan floods (IASC, 2011a) and Haiti earthquake (IASC, 2011b). This has been accompanied by the development of inter-agency agreed RTE procedures (IASC, 2010). The Emergency Capacity Building (ECB) project developed guidance on joint evaluation (2011) and conducted a series of country-based and regional-level workshops to strengthen ECB member capacities to commission and carry out impact assessments (ECB-AIM Standing Team, 2011; ECB, 2007). Most recently, Oxfam has developed some evaluation guidance around contribution to change and humanitarian programme impacts using a livelihood-based approach (2013, forthcoming).

Moreover, a recent General Assembly resolution emphasises the importance of an effective evaluation function for organisations of the UN system, and of promoting a culture of evaluation. The resolution calls upon members of the UN aid system to ensure the effective utilisation of findings, recommendations and lessons learned (UN General Assembly, 2012). Also notably, UNEG and the OECD-DAC group have been working on improving approaches to carrying out peer-reviews of their members' evaluation systems and structures (Liverani and Lundgren, 2007; UNEG, 2011; 2012d).

Approach to the study

Many of the challenges and complexities confronting the humanitarian endeavour – especially those relating to the scale and inter-connectedness of its ambition, and to the difficulties of understanding and attributing outcomes and impacts – are mirrored by the challenges confronting evaluation work. However, the feeling that the full potential of evaluation to bring about change in humanitarian practice is not being realised comes down to more than just the complexity of the system. There are many other reasons – discussion of these provides the content of this paper. This discussion is enriched by suggestions of possible solutions used by humanitarian agencies to address these challenges.

The publication of this paper study represents the conclusion of the second and final phase of a stream of ALNAP work looking into issues related to evaluation capacities. As noted above, the trigger to this study was the publication of the 2006 ALNAP study on evaluation utilisation in the humanitarian sector. The paper suggested that more work would be needed to improve understanding of the evaluation function and evaluation use in humanitarian practice (Sandison, 2006: 90, 139-141).

Building on the interest shown by ALNAP members to explore these issues further, ALNAP commissioned a follow-up work, to scope the subject further, and get a better sense from members of the key issues for discussion. That first phase of the research culminated in a working paper (Hallam, 2011), which built upon ideas generated through a literature review, interviews with humanitarian evaluators and consumers of evaluations as well as a workshop held in September 2010 that brought together evaluators from UN, NGO, Red Cross and other donor organisations, as well as independent evaluation consultants and academics. The workshop was also used to consult the ALNAP membership about: (i) narrowing the focus of subsequent research; and (ii) exploring which elements may contribute or hinder individual, team and organisational capacities to fund, commission, deliver and use humanitarian evaluation more effectively.

That initial working paper proposed a draft framework of evaluation capacities (Hallam, 2011: 5-6). This was designed to assist ALNAP members to strengthen their evaluation efforts by helping them 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. In its draft form, the framework was also intended to provide a canvas on which to plot and situate the inputs collected between 2011 and 2013, through group consultations and one-on-one conversations with evaluation officers, humanitarian evaluation consultants and programme staff working closely with the commission or use of evaluation[1].

Phase two of this research has involved a series of workshops held in London, Geneva, Washington, Madrid and Tokyo, where the initial working paper was discussed and a more nuanced understanding of the challenges and opportunities related to evaluation capacities gradually emerged. Ideas generated at these meetings related to promoting utilisation of evaluation in the widest sense and improving evaluation capacities. For this final paper, further interviews have been carried out as well as a new search of

[1]Between 2010 and 2013, ALNAP facilitated seven workshops in five countries. These attracted more than 115 participants from over 40 organisations. Workshop materials are available at: http://www.alnap.org/using-evaluation

published and grey literature covering both EHA and development evaluation literature. ALNAP also designed and launched a community of practice (ALNAP Evaluation CoP) with the objective of soliciting and garnering suggestions, comments and experience from evaluators within the Network (see Box 1).

Since its inception, this research has been collaborative, and hinged on a blend of desk-based and more consultative and participatory activities. This blend allowed evaluators from different agencies in the ALNAP Network to harness concrete examples and, we hope, some untold evaluation stories and evaluator insights that rarely find space to be documented in more formal reviews and evaluation reports. A list of contributors is provided at page 85.

In summary, following the first working paper on the same subject (Hallam, 2011), this study is the result of:

a. a second round of group consultations with evaluators
b. a second review of evaluation literature covering both humanitarian-specific as well as broader development and programme evaluation literature
c. one-to-one interviews with different agency staff engaging more or less directly with evaluation work, representing the different constituencies within ALNAP: UN agencies, Red Cross and Red Crescent (RC/RC) movement, donor organisations, NGOs, independent consultants and academics
d. a virtual conversation with ALNAP members over one year (June 2012 to May 2013) on the Evaluation CoP, where different evaluator insights have been discussed, expanded, confuted or validated by different contributors.

Promisingly, many agencies and organisations now seem to be paying much more attention to evaluation capacities and evaluation utilisation in its broadest sense. Their endeavours and ideas have enriched this paper immeasurably.

Untold evaluation stories and evaluator insights that rarely find space to be documented.

Box 1

A community of practice for consultation and peer-learning among evaluators

In June 2012, ALNAP established a face-to-face and online Evaluation Community of Practice (CoP) with a twofold objective:

1. To facilitate a consultation with evaluators in the Network on the relevance, clarity and usefulness of a draft framework on humanitarian evaluation capacities proposed in an earlier ALNAP Working Paper (Hallam, 2011).
2. To provide an opportunity for evaluation professionals within ALNAP to come together – in person and/or virtually – to ask questions and share learning and to get practical advice from colleagues and peers in similar situations.

Workshops were hosted by various ALNAP Members to discuss issues relating to organisational support to evaluation, leadership, culture, communication, dissemination and uptake of evaluation have been a strong focus. These peer-to-peer learning meetings attracted staff in programme management, planning, coordination, donor relations and policy advisory roles, as well as evaluators.[2]

[2] Workshop reports and other materials from the consultations and peer-learning events are available at www.alnap.org/using-evaluation

Objectives of the study

The main objective of this paper is to motivate and encourage humanitarian evaluators, by highlighting and discussing concrete ways to address the challenge of poor or ineffective use of evaluation. The use of insights and case studies from evaluation colleagues in the humanitarian sector is designed to reinforce the sense that much is possible. It is hoped that this study will contribute to strengthening the capacities of individuals, teams and organisations to become better-informed commissioners, users and consumers of evaluative information and knowledge. The three main approaches to this are to:

- highlight the key issues around improving use of evaluation of humanitarian action
- present a framework for analysing and some practical suggestions for improving the capacity of agencies to fund, commission, support, carry out, and meaningfully use humanitarian evaluations
- further illuminate the issues by providing case studies and insights from evaluators that explore what has worked in practice.

The primary audience of this report is humanitarian agency staff. There are publications on capacity building in the development sphere, but very little that speaks directly to the humanitarian audience. This report seeks to address this gap. In these straitened economic times, the focus should not necessarily be on spending less and doing more, but on a set of strategies that all those engaging with humanitarian evaluation work should be able to relate to, and draw upon to increase the effectiveness of their work: ranging from the more fundamental to the 'easy-wins'.

When gathering case studies and evaluator insights, it often proved harder than expected to fit them squarely into single sections of the framework. In the end, and as should have been expected, we found that good evaluation depends on doing lots of things right, rather than necessarily excelling in a single area of the framework.

Analytical framework for improving the understanding and use of EHA

Undertaking evaluation work and ensuring its quality is worthwhile if the activity leads to some use of the findings, and contributes to improving knowledge among those best placed to use it and bring about change and improvements in practice (Weiss, 1990; Feinstein, 2002; Grasso, 2003; Tavistock Institute, 2003: 79; Preskill and Boyle, 2008; Rist et al., 2013). If we wish to ensure that evaluations contribute to change and improvement, it is worth exploring how these changes are meant to happen. It is often assumed that evaluations yield information or provide lessons that flow directly back into the policy cycle and are thus incorporated in the planning of future programmes and projects. In this way, there should be a constant learning process leading to ever-improving performance (Preskill and Torres, 1999; Frerks and Hilhorst, 2002; Ramalingam, 2011). This understanding presupposes a rational, scientific planning model. However, 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 (Weiss ibid., 1990; Pawson and Tilley, 1997; Frerks and Hilhorst ibid., 2002; Start and Hovland, 2004; Rist et al ibid., 2013). Box 2 outlines different ways in which evaluations can be used in practice.

Box 2

A menu of evaluation uses

A 2006 ALNAP paper on evaluation use (Sandison, 2006: 92-97) highlighted strengths and weaknesses in how humanitarian organisations were moving towards utilisation and a user-focused approach to EHA. The study discussed the various ways in which an evaluation can be used and misused. It highlighted much of the 'classic' debate on evaluation utilisation from programme evaluation literature, as follows.

- **Direct, or instrumental use:** single evaluations may be used directly or in an 'instrumental' manner whereby the results, findings, conclusions and recommendations are taken up. In practice this is unusual and where it does occur it tends to take place only partially.
- **Conceptual use:** more often, several evaluations or individual evaluations combined with other evidence and opinion are used cumulatively to inform debates and influence decision-making.
- **Process use:** even where evaluation results are not directly used, the process of carrying out an evaluation can offer opportunities to exchange information and clarify thinking. 'Process use' occurs when those involved learn from the evaluation process or make changes based on this process rather than just on the evaluation findings (Patton, 2010a: 145).
- **Symbolic use:** evaluations can be used to justify pre-existing preferences and actions. Many evaluation conclusions affirm what people already know about perceived successes and shortfalls of a programme. The use of such knowledge is not dramatic, or even visible, but it bolsters the confidence of those who want to press for change (Weiss, 1998: 317) or justify previously made decisions or for advocacy and marketing (McNulty, 2012).

Change is not a rational process, but one that occurs through a complex variety of interactions across an organisation. As a result, harnessing the full power of EHA for change requires attention to a whole range of inter-connected issues.

These issues involve much more than simply 'using' evaluations better, though this is of course desirable. 'Evaluative thinking' should be rewarded, and should permeate throughout the organisation. Yet, people must be willing to change, and be rewarded for doing so, for improvements to occur in their programmes. They need to have the tools and the time to analyse what they do, and to explore the findings in a way that will encourage change. The organisation as a whole needs to accept that this will involve challenge to existing ways of working and critical debate. Leadership and culture are of great importance, as are formal and informal systems and reward structures.

To facilitate analysis and discussion of the factors that influence whether evaluations lead to impact, we have grouped our observations into an analytical framework. The purpose of this is to help shape discussions and stimulate thought and action on improving the understanding and use of EHA. The framework draws on the literature on evaluation capacity development as well as on evaluation use, and was influenced by discussions with humanitarian evaluators as well as by conversations within the ALNAP Evaluation Community of Practice (Box 1). The framework has been designed to be as light and as easy to use as possible, with a clear focus on those issues most pertinent to humanitarian evaluation.

We present the outline of the framework here, with more detailed discussion of each capacity area in the three Sections that follow. Case study material and evaluator insights are inserted into the framework where appropriate. We look first at overarching issues of leadership, culture and resource (Section 1), followed by issues of purpose, demand and strategy (Section 2). Next are more specific evaluation processes and systems (Section 3).

CAPACITY AREA 1

LEADERSHIP, CULTURE, STRUCTURE AND RESOURCES

Ensure leadership is supportive of evaluation and monitoring

Interviews carried out for this study identified leadership as key in improving the impact of EHA. The same finding appears in both the academic literature (Weiss, 1990; Torres, 1991; Mayne and Rist, 2006; Heider, 2011) and reviews of the evaluation functions of humanitarian organisations (Foresti, 2007; UNEG, 2012a). 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 discussing performance issues, there may be some reluctance in accepting evaluation findings and constraints in 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. Modelled leadership behaviour is often a lynchpin to fostering an enabling environment for evaluation and its subsequent utilisation (Box 3) (Buchanan-Smith with Scriven, 2011; Heider, 2011; Tavistock Institute, 2003; Clarke and Ramalingam, 2008; Beck, 2004b).

Box 3

The importance of leadership in evaluation use

Organizational leadership that supports evaluation and makes deliberate use of evaluation findings is critical to the functioning of an internal evaluation office. Organizational leaders need to persuade staff members that evaluation is part of organizational life in order to gain their support for evaluation activities and build trust in evaluation findings (Boyne et al., 2004). Further, organizational leaders need to identify self-reflection and program improvement as key management principles espoused by the organization and encourage staff to engage in these activities through evaluative inquiry (Minnett, 1999). In order for this to happen, the entire organization should be knowledgeable about evaluation (Boyne et al., ibid.) and should feel ownership in the evaluation process (Bourgeois et al., 2011: 231).

What can evaluation departments do to improve leadership support for evaluation and monitoring? Here are a few suggestions that are illustrated in more detail in evaluator insights throughout the remainder of the text.

- Identify an evaluation champion who can show the wider organisation the benefits that can arise from evaluation.

- Draw attention to good-quality evaluations that highlight important issues within programmes that would otherwise have been missed, as these can serve as a tool to convince managers of the importance of evaluation.
- Seek to ensure that an organisation's recruitment policy for senior managers, and orientation sessions, emphasise the importance of evaluation and its use as a tool for learning and reflection.
- Appoint senior staff with specific responsibility for evaluation and knowledge management.
- Make sure that the organisation's evaluation department focuses on information needs of senior management, to keep them engaged and supportive.

Evaluator insight 1

Focusing evaluation on questions coming from senior leadership

UNRWA has recently been undergoing a multi-year period of transformation, and this has given the evaluation department an opportunity to identify questions of particular use to leaders in the organisation responsible for making strategic decisions. Evaluators work with the management committee to identify questions relevant for specific departments at that particular phase of the process, and check with other key stakeholders, such as donors and hosts, to ensure that they also feel the questions are pertinent. This allows the prioritisation of evaluations which make a real contribution to the direction of the organisation.

Source: UNRWA Official, personal communication at UNEG meeting, April 2013.

A study on strengthening learning from research and evaluation in DFID (Jones and Mendizabal, 2010) suggested the appointment of a director as 'knowledge and learning' champion. Along these lines, UNICEF has created a dedicated post focusing on humanitarian knowledge management and learning capacity within its emergency department. UNICEF has also recently carried out a review of its evaluations of humanitarian action (2013). This made the following recommendations.

- EHA orientation shall be included as a session in the orientation programme for Representatives and Deputy Representatives by the end of 2013.
- EHA awareness will be included in Regional meetings beginning in 2013 and going forward (e.g. Regional Management meetings; Deputy Representatives and Operations meetings; Regional Emergency meetings).

Evaluator insight 2

Change in management - a window of opportunity for evaluation?

Evaluation units or teams should be able and willing to take advantage of windows of opportunity to catalyse discussions on evaluation, learning and improvement. A change in senior management may provide such a window. For example, in the case of one UN agency, weaknesses in its humanitarian response were recurrently highlighted in evaluations spanning over a decade, but these received little attention in terms of follow-up. However, the re-emergence of these challenges during the response to the Haiti earthquake happened to coincide with a change in senior management. This provided an opportunity for a fresh look.

The new Executive Director was keen to learn and engage from the very beginning with an evaluation commissioned in Haiti. The evaluation unit used previous knowledge to enrich the conversation and brought forward concrete analysis that was packaged appropriately for the new management to act upon. Senior leadership engagement was vital in keeping the review relevant to major policy currents in the organisation, and sustaining attention to and positive engagement with the review. Furthermore, involvement of the Director's office helped secure a swift management response that led to major changes to improve the organisation's performance in large-scale emergencies.

These changes included: simplified standard operating procedures; greater integration of cluster work within its trainings and guidance for senior managers; fine-tuned human-resources processes for getting the right people on the ground at the right time; and strengthened guidance on responding to urban disasters.

Source: Abhijit Bhattacharjee, Independent, personal communication, June 2013.

Evaluator insight 3

Leadership support to evaluation – an example from the Global Education Cluster evaluation

UNICEF and Save the Children carried out an evaluative review of the Global Education Cluster. This review was much appreciated and well used. A major reason for this success was that, because of the importance of the cluster approach to both organisations, leadership on both sides was very supportive of the evaluation even though both knew that the review would include criticisms of them. Following the review, a joint management response was subsequently issued: cluster objectives and indicators were clarified, as were roles and responsibilities; there was joint planning, budgeting and resource mobilisation; and stronger governance arrangements were designed and implemented. The review was fairly critical of UNICEF in places but this did not prevent a very good management response from them and active follow-up since completion.

Source: Rob McCouch, UNICEF, personal communication, October 2012.

Promote an evaluation culture

Enabling credible evaluation that is useful and gets used requires establishing effective systems and structures that support evaluation work through budgeting, human-resources allocation, and appropriate knowledge management and learning processes (Rist et al., 2011; Heider 2011). Both Patton (1997) 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 are likely to end up as burdensome and potentially counter-productive. Where the culture is conducive, an organisation is more likely to actively seek information on performance, including evaluation data, to learn and to change its behaviour.

Working towards fostering and sustaining an evaluation culture is not an end in itself. But it can become an integral part of moving towards becoming a learning-oriented organisation (Guba and Lincoln, 1989; Weiss, 1998; Preskill and Torres, 1999; Mayne and Rist, 2006; Mayne, 2008; Patton, 2010a; Heider, 2011). This move is supported when fostering an evaluation culture entails:

- stimulating the capacity of staff to demand, generate and use high-quality evaluation products
- making space for staff and programmes to discuss programme quality and performance, including critical elements and failure
- appreciating the range of purposes and uses of evaluation
- framing evaluation more within learning and improvement, and less within top-down, mandated accountability
- stimulating the demand for evaluative knowledge among staff in the organisation at different levels (from programme to policy departments)
- expecting programmes and interventions to be planned, designed and monitoring in a manner conducive to subsequent evaluation
- recognising the limits of evaluation, and the need to combine quantitative and qualitative evidence to support evaluation findings and conclusions
- acknowledging the different needs of different users of evaluation[3].

At the broadest level, a culture of learning and accountability (both *upwards* towards institutions and funders, as well as *forward* to crisis-affected communities) can reinforce the enabling environment necessary for evaluation to bring about change and improvements in humanitarian practice. The 'enabling environment for evaluation' has been described as the degree to which information about past performance is sought, and the extent to which there is a drive to continuously reflect, learn, improve and hold people responsible for actions taken, resources spent, and results achieved (Heider, 2011: 89). Such a culture is also embedded in tacit norms of behaviour including the understanding of what can and should (and should not) be done to address shortcomings and strengthen positive behaviour modelling (Tavistock Institute, 2003; Heider, 2011).

[3] Indeed, many of the points above echo some of the core tenets in the literature on evaluation utilisation (see for instance Chelimsky, 1997; Weiss, 1998; Feinstein, 2002; Patton, 2008).

A culture of learning and accountability can reinforce the enabling environment necessary for evaluation to bring about change and improvements in humanitarian practice.

So, how do organisations go about cultivating a positive internal culture? If it is true that success breeds success, then high-quality evaluations that address real information needs, shall also improve the way in which evaluation work in itself is regarded. This can positively affect the evaluation culture, which, in turn, increases the demand for evaluations, and makes it more likely that evaluation systems and structures are improved sustainably. A virtuous circle ensues. Tackling issues at each stage of the framework identified in this paper can help in creating that virtuous circle. However, each organisation also 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 may 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.

Despite the differences between organisations, some general themes have emerged through this study on to how to improve an evaluation culture:

- decrease perception of evaluation as criticism and evaluators as critics
- use both internal and external personnel to carry out evaluations
- re-brand evaluations
- be flexible and have fun
- get incentives right.

We discuss these themes in the following pages.

Those promoting change may need to be opportunistic, seizing the 'easy wins' first, so that impact is quickly seen.

Decrease perception of evaluation as criticism and evaluators as critics

Evaluation may generate apprehension in those engaging with, or subject to it. Some may see evaluators as critics, resulting in resistance to their recommendations for change (see for instance Weiss, 1998: Chapters 2 and 5). Several recent conversations with ALNAP members highlight a variety of approaches for changing these perceptions.

Evaluator insight 4

Oxfam Intermón made criticisms less personal

The organisation used an ALNAP paper on lessons from flood responses[4] as a key background document in evaluating its flood-response programme in Mozambique. The team in Oxfam Intermón was able to link many of the findings from their own evaluation to those detailed in the ALNAP paper. This allowed the team to see that certain problems and challenges were common to other responses, and experienced also by other agencies in other countries, often for systemic reasons. This changed how those in the country team perceived the evaluation, and made it less threatening and more useful.

Source: Luz Gómez Saavedra, Oxfam Intermón, personal communication, March 2013.

[4] Alam, K. (2008) 'Flood disasters: learning from previous relief and recovery operations'. London: ALNAP/Provention consortium (www.alnap.org/pool/files/ALNAP-ProVention_flood_lessons.pdf).

Evaluator insight 5

Groupe URD adopted an 'iterative evaluation process'

This helps create stronger relationships between members of staff, external evaluators, and the organisation. The evaluators gain a better understanding of the organisation and its operating context. This approach also helps to reduce the perception that evaluators are 'parachuted in'.

Groupe URD conducted three or four evaluations of the same project over a period of two years. The later evaluations concentrate largely on identifying progress made with the recommendations of the previous mission and identification of new challenges. 'This means that the evaluators have to really take ownership of the recommendations – they can't just drop them and walk away – because they will be going back to see how they are working, and whether they have had an impact.'

This leads to a powerful dialogue between the evaluator and the programme staff that goes on over the life of the project. [...] The cost of this approach is that the evaluator loses a degree of their independence (although hopefully not their objectivity) in order to become an agent of change. But [...] the gains in improvement – which is, after all, the main purpose – make this worthwhile.

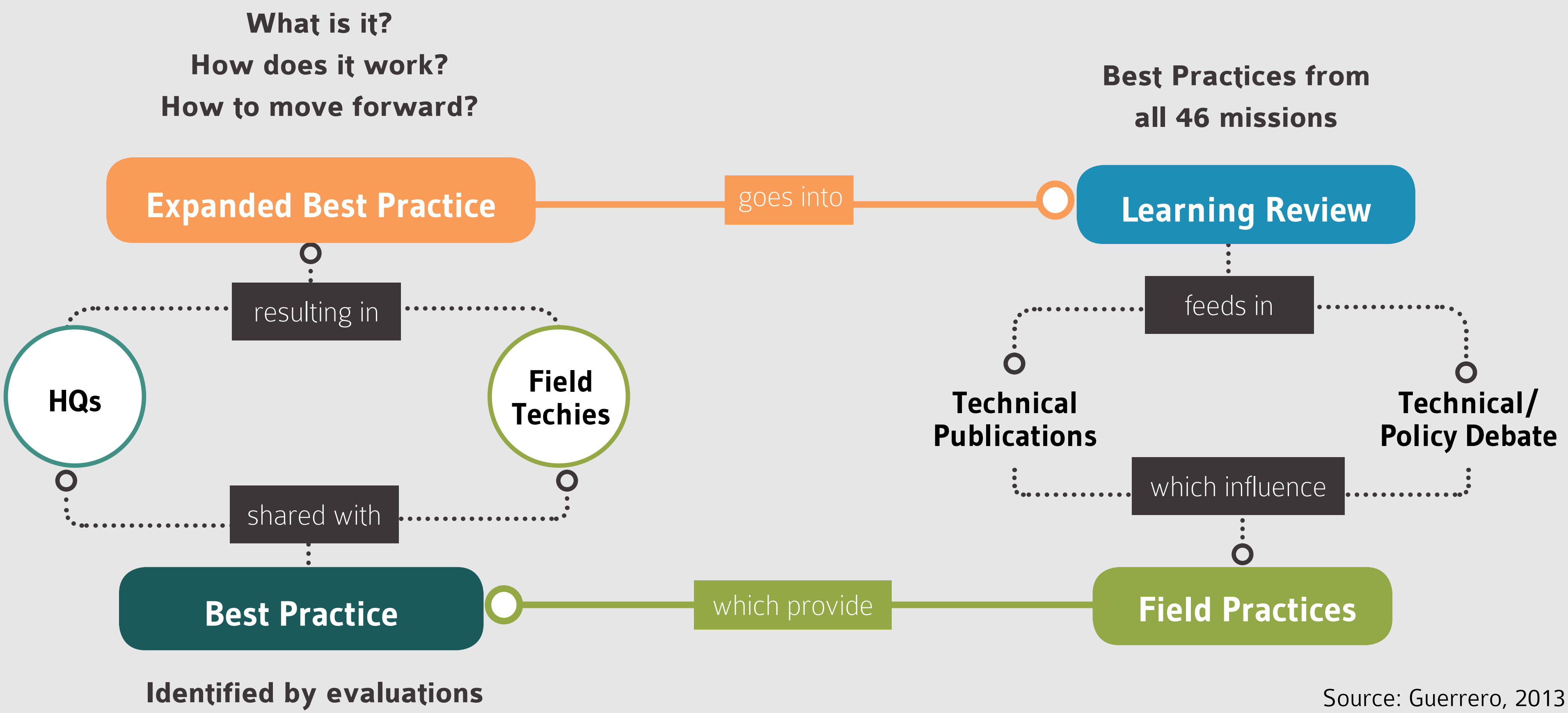
Source: Francois Grünewald, Groupe URD, ALNAP Evaluation CoP contribution, November 2012

Other organisations emphasise positive findings and good practice examples in their evaluations. A strong example of this is from Action Against Hunger (ACF). Part of ACF's new Evaluation Policy and Guideline, is systematically asking evaluators to identify 'one particular aspect that shows strength for each ACF project evaluated' (ACF 2011a). This practice intends to 'not exhaustively present a programme approach but to showcase a strong element, from which other ACF and humanitarian/development professionals can draw inspiration and benefit' (ACF, 2013: 31). A selection of 'Best Practices' that complete the required cross-operational-level loop (Box 4), which consists of steps to validate and elaborate the selected aspect, are then published in the annual Learning Review (ACF, 2011b; 2013).

Box 4

Action Against Hunger's process for garnering and validating best practices

Recent ACF practice in garnering, elaborating and validating emerging good practice involves following a series of steps as depicted in the learning loop below (starting from the bottom left-hand corner).



Source: Guerrero, 2013.

ACF personnel feel that positive results from this exercise are already being observed. For instance, this practice has contributed to building better relationships with evaluators as communication is kept open – after the completion of the evaluation report and deliverables – and they can see how their initial recommendations have evolved and been taken up by the organisation. ACF has also started seeing learning come full circle at different levels of the organisation. For instance, 'Best Practices' are spreading organically throughout the organisation; with the field-based teams taking the lead on their adoption. Encouragingly, the ACF Evaluation and Learning Unit has also started noticing how new programme and policy development are more explicitly making reference to the evaluative evidence produced (Guerrero, 2013).

Of course, criticism cannot always be avoided. Some NGOs opt for internal rather than external evaluations, judging that 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. Honest and frank discussion may be constrained where evaluations are to be made available to the public.

Use both internal and external personnel to carry out evaluations

A review of the literature and discussions with field staff reveals a common theme about field and operational staff being unhappy with current evaluation practices. Ideally, outsiders would be required to learn about the culture and practice of the organisation being evaluated. Many field-based personnel feel that this requires a significant investment of their time if an evaluation is to lead to more pertinent and actionable recommendations. When this does not happen, this shortcoming can damage the credibility of the evaluation and feed into the often-held perception that nothing seems to change as a result of evaluation work[5].

During this study, many interviewees mentioned that they have started using more internal evaluation, or mixed teams of insiders and outsiders. Recent literature explains that one of the main benefits of using internal evaluators is their greater ability to facilitate evaluation use. Compared to external consultants, insiders are 'more likely to have a better understanding of the concerns of field personnel, and of their perspective on key issues. Thus, they are more likely to find means to support evaluation utilisation

[5]For a discussion on this point, see Patton (1990).

by turning findings into appropriate actionable items, and are better able to follow through on their recommendations through monitoring activities' (Bourgeois et al., 2011: 230). Apart from supporting greater direct or instrumental use of evaluation findings, internal evaluators are more likely also to foster process use within their organisation and to support organisational learning at a higher level[6].

Despite these benefits however, some authors suggest that 'hybrid models' (combining both internal and external evaluation staff) may often be the most appropriate (Bourgeois et al., 2011). Some organisations insist on external evaluators to maintain independence and to ensure that radical recommendations are not suppressed[7]. Yet, by involving internal staff in evaluation work, evaluation units can also:

1. play an important role in knowledge transfer and dissemination
2. learn about the most relevant issues to cover in future studies and evaluations (Weiss, 1998: chapter 2; Minnett, 1999)
3. play a role in maintaining organisational memory and ensuring that important lessons are not lost with time (Owen and Lambert, 1995).

Evaluator insight 6

UNHCR's experience with mixed evaluation teams

The evaluation unit at UNHCR has in the past experimented with using fully independent evaluation teams. However, more recently, this practice has changed, with the unit now favouring more mixed teams, involving both members of the evaluation unit and external consultants. 'Independent teams can be very expensive to hire, difficult to manage and at the end of the day take much of the learning with them', remarked the Head of Unit who also argued that, in his position, 'it would be very difficult to attract the best and brightest staff to the evaluation unit if all that was on offer there was the task of managing consultants. Bringing in people from elsewhere within the organisation to the evaluation unit means that field experience and insights are regularly refreshed and reflecting evolving practice of UNHCR operations on the ground.'

Source: Jeff Crisp, UNHCR, personal communication, April 2012.

[6]Also see UNEG (2010) and sections 2.2 to 2.5 and 3.9 of Cosgrave and Buchanan-Smith (2013).

[7]Tony Beck and John Cosgrave, Independents, personal communications, May and June 2013 respectively.

CARE is another organisation using mixed evaluation teams:

> *an evaluation of tropical storm Jeanne was particularly effective partly because, in addition to external consultants, CARE Haiti's monitoring and evaluation (M&E) focal point also joined the evaluation team, helping to ensure recommendations were realistic and followed-up. The CARE staff member himself learned a great deal. The report was also translated into French, which improved its use amongst local staff and partner agencies. (Oliver, 2007: 14)*

Re-brand evaluations

Putting aside the time to learn and reflect can be very difficult at all levels within a busy humanitarian organisation. However, evaluation faces specific challenges, some of which are about the way they are now regarded, perhaps because of a failure to use them appropriately in the past. (In essence, once something is called an 'evaluation' people don't want it!) Re-branding evaluation, as part of a change in how evaluations are conducted and used, can help.

In this vein, some interviewees went as far to suggest that the very term 'evaluator' be swapped for 'facilitator'[8].

> *Where evaluation is used for learning and adaptive management, this implies a fundamental change in the role of the evaluator. From being a remote, research-oriented person trying to systematise the known and unearth the hidden, he or she becomes a process facilitator. His or her role is to help design and organise others' inquiry and learning effectively. Stakeholder analysis and communication skills, as well as the ability to manage group dynamics, become prime assets. [...] Enhancing learning will eventually mean a fundamental restructuring of the training, methodological baggage, professional skills and outlook of evaluators. (Engel et al., 2003: 5)*

[8]This approach echoes some of the most recent evaluation thinking reflected in Patton's developmental evaluation approach (Patton, 2010a).

Furthermore, some organisations use evaluation as a means to promote and celebrate their work. For example, the IFRC Secretariat's evaluation framework states that:

> *Reliable evaluations can be used for resource mobilisation, advocacy, and to recognise and celebrate our accomplishments. The promotion of a programme or policy through evaluation is not perceived as a pure marketing tactic because evaluations provide impartial and often independent assessments of our performance and results, lending credibility to our achievements. They help demonstrate the returns we get from the investment of resources, and celebrate our hard effort.*
> *(IFRC, 2011a: 2)*

Additionally, as illustrated in one CARE evaluation, highlighting what they had done well, rather than remaining limited to where their response effort had fallen short, can boost staff morale and increase its effectiveness (Oliver, 2007).

Similarly, a 2009 World Bank paper looking at use of impact evaluations reported:

> *There is always demand for results that confirm what people want to hear. Concerns over potential negative results, bad publicity, or improper handling of the results may reduce demand; sensitivity, trust-building, and creative arrangements may help overcome these fears. Consequently, there may be some benefit in taking advantage of opportunities to present good results, especially if it helps the process of getting stakeholders to understand and appreciate the role of impact evaluation.*
> *(World Bank, 2009: 68)*

Finally, a number of organisations produce annual evaluation reports. Some of these – such as WFP's annual evaluation reports[9], or CIDA's Lessons learned report (CIDA, 2012) are well-presented and content-rich summaries of key evaluation findings. These may help refresh the image of the evaluation unit by reducing the perception that evaluation units work only on lengthy evaluation

[9]The repository is available at: www.wfp.org/about/evaluation/key-documents/annual-evaluation-reports

reports specific to one situation or theme. This may help to bring evaluation closer to other members of the organisation. Further, these summary reports can help to show managers and others what can be achieved through evaluation.

Be flexible and have fun

One approach to improving evaluation culture is to add some flexibility, fun and creativity to the process of evaluation. One agency evaluation head reports a personal 'obsession' with fighting boring evaluation reports:

> *Evaluators often go to incredibly interesting places, meet a wide range of amazing people but then proceed to write reports which are totally devoid of interest and flair. We all have far more on our desk than we can hope to read. Long and poorly written reports that are stuffed with jargon and annexes are not likely to attract the attention of key stakeholders.*

Suggestions for improving reports include the use of quotations, anecdotes (when they make a telling point), short paragraphs and snappy titles to attract the reader's attention.

Evaluator insight 7

It is no crime to be flexible

There is a lot to be said for adopting a flexible approach to terms of reference (ToRs), amending them as necessary in the course of an evaluation project, to take account of new and different issues not anticipated when an evaluation project was launched. If something important or interesting arises in the course of an evaluation then it would be crazy not to revise the terms of reference accordingly. Terms of reference do not have to be structured around the classical evaluation criteria (effectiveness, efficiency, coverage, coherence, etc.). While such concepts may well be used in an evaluation report, the 'art' of evaluation requires us to avoid formulaic approaches.

Source: Jeff Crisp, UNHCR, ALNAP Evaluation CoP contribution, July 2012.

A CARE study of the use of evaluation found that 'the overwhelming sentiment regarding evaluation reports was that they are too long and too tedious to sift through' (Oliver, 2007: 2-3).

Evaluator insight 8

Flexibility in evaluation design

I believe that flexibility in designing and conducting evaluations has been our key factor for success. Sometimes quick lessons-learned exercises have been more useful than comprehensive evaluation processes. Occasionally it has been rewarding to take time for a transversal/thematic evaluation and allow for thinking and reflecting about (possibly) surprising findings through several intermediate debriefings with stakeholders. We have also learned to adjust the timing for project evaluations to the annual planning circle, whereby we ensure that results are directly integrated in the next planning review.

Source: Sabine Kampmueller, MSF, ALNAP Evaluation CoP contribution, October 2012.

Get incentives right

USAID's 2009 synthesis of evaluation trends concludes that the link between evaluation and action is a function of organisational culture and incentives:

> *Structures for systematically processing evaluations, rather than shelving those that deliver unsettling results, are a critical element in evaluation cultures that are likely to act on evaluation findings.... Problems arise when incentives within agencies are not consistent with a learning model and when staff commitment to evaluation is uneven across levels of the organizational hierarchy (USAID, 2009: 27).*

Jones et al. (2009) found that, in DFID, staff avoided learning from closed or failed/failing projects or programmes, because of misaligned learning and career incentives.

Discussing evaluation incentives, a recent paper commissioned by InterAction made the following suggestions:

- *Conduct an internal review of current incentives and disincentives for learning. Once you know the ways in which learning is discouraged (for example, in how failure is perceived and penalised), it is possible to restart a culture-building process with a bonfire of learning disincentives.*
- *Start giving rewards for learning achievements – as opposed to 'coming out well' in an evaluation – in the form of recognition, and even remuneration.*
- *Build the use of evaluation into job descriptions and performance appraisals. Anyone who commissions or contributes to an evaluation should be held accountable to see that findings are*

utilised. 'Accountability' (in this context) must be about creating the conditions for frank and open discussions of results. Accountability is for enabling learning, not for getting 'good' evaluation results. (Bonbright, 2012).

Even the best plans and practices are unlikely to sustain rigorous evaluation use over time unless internal and external incentives reinforce effective use of evaluations. Mackay (2009) discusses three types of incentive: carrots, sticks and sermons, and reports that many of these incentives help to institutionalise monitoring and evaluation (M&E) in developed and developing country governments. Carrots provide positive encouragement and rewards for conducting M&E and utilising the findings. Sticks include prods or penalties for those who fail to take performance and M&E seriously. Sermons include high-level statements of endorsement and advocacy concerning the importance of M&E. Although written with country-led M&E systems in mind, a selection of these ideas more relevant to EHA is presented below, adapted slightly for the humanitarian context.

Carrots

- Awards or prizes – high-level recognition of good or best-practice evaluations.
- Provision of additional funding to conduct M&E.
- Assistance to programme areas in conduct of M&E – via help-desk advice, manuals, free training, etc. This makes it easier to carry out M&E and to use the findings.

Sticks

- Enact policies which mandate the planning, conduct, and reporting of M&E.
- Withhold part of funding if M&E is not done.
- Penalise non-compliance with agreed evaluation recommendations.

Sermons

- High-level statements of endorsement by chief executives and other senior managers.
- Awareness-raising seminars/workshops to demystify M&E provide comfort about its feasibility, and to explain what's in it for participants.
- Use actual examples of influential M&E to demonstrate its utility and cost-effectiveness.

Even the best plans and practices are unlikely to sustain rigorous evaluation use over time unless internal and external incentives reinforce effective use of evaluations.

The importance of getting incentives right is highlighted by Jodi Nelson, Director of Strategy, Measurement, and Evaluation at the Bill & Melinda Gates Foundation:

> *The biggest learning for me is that my job is more about organizational change than it is about being an evaluation expert. An organization can have great M&E people and expertise, as we do, but it won't actually lead to anything unless there's alignment up and down the organization around what enables success. Some examples include: leaders that continually ask their teams to define and plan toward measurable outcomes, consistent expectations for staff and partners about what constitutes credible evidence for decision making, executive leaders who understand and sponsor change that can be tough and take a long time, and tools and resources for staff and grantees to integrate rigorous planning and M&E into their day-to-day work.* (Nelson, 2012)

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, staff's capacity to use evidence and lessons from research and evaluations helps build 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, Jones and Mendizabal (2010) 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', while 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 informed by research and evaluation, to support their arguments.

Create organisational structures that promote evaluation

Evaluation units have various locations and reporting lines within agencies (Foresti, 2007)[10]. Some organisations have a separate and/or central evaluation unit responsible for carrying out evaluations. In others, responsibility is decentralised throughout the organisation (Hallam, 1998). The type of management structure potentially 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 – by putting evaluation in a 'silo' – could undermine broader institutional learning.

Some evaluation departments report directly to the Board, or to the Chief Executive. Other evaluation departments are lower in the hierarchy. A study commissioned by the Agence Française de Développement (Foresti, 2007) concluded that there is no right or wrong approach here, but that the position of the department has an impact on learning and accountability as well as on the perceived independence of the evaluation unit. Furthermore, the location of the evaluation function is likely to have an influence on

[10]This approach echoes some of the most recent evaluation thinking reflected in Patton's developmental evaluation approach (Patton, 2010a).

its accountability to senior management, and how much resistance it may encounter from programmes and staff. The same study also argued that there is generally 'a tension between the independence of evaluation departments and their success in engaging users of evaluation' (Foresti, 2007: 8). It is necessary to balance this independence and other issues such as credibility and utilisation. Too much independence may lead to lack of utilisation, while too little independence may reduce credibility[11].

"My job is more about organizational change than it is about being an evaluation expert."

Evaluator insight 9

Tearfund example of integrating evaluation across organisation functions

Tearfund's monitoring and evaluation (M&E) function is based within the Learning Information Advice and Support Team (part of the People and Organisational Development Group), and reports to the Head of Knowledge Management. The team works across the organisation, maintaining strong working relationships with the International Group and the Head of Strategy, as well as donor and other supporter relations teams.

This location within Tearfund appears to be conducive to the cross-fertilisation of learning and ideas. The M&E function brokers across teams and groups, and is conducive to raising the bar on the undertaking and utilisation of evaluations.

Source: Catriona Dejean, Tearfund, personal communication, June 2013.

What works for one organisation might not work for another. So, in a small NGO, it might be much better for the evaluation office to be close to the programmes and less structurally independent. For a large UN organisation, where there may be huge differences in power and perception between the layers of the organisation, a more structurally independent evaluation office may be required to be able to 'tell truth to power'.

[11] John Cosgrave, Independent, personal communication, June 2013.

Regardless of the size of an organisation and its reporting structures, a minimum level of independence is required, and the evaluation team should not be subject to pressure and competition that may arise from operational departments. (Heider, 2011). The location of the evaluation department or team should be conducive to the reduction of barriers to evaluation use[12].

Interestingly, we found no case studies demonstrating how a change in an organisation's structure had led to a change in the use and impact of EHA. Perhaps there is a gap here in the research or perhaps this shows that, although we might expect structure to be very important, other issues (such as culture and leadership, and evaluation processes) are more important.

Secure adequate resources – financial and human

Many field workers and managers complain about the costs of evaluations. However, they do not always stop to consider the potentially huge financial and human costs of continuing with unsuccessful programmes. Research into the costs and benefits of evaluation could lead to a better understanding of the concrete benefits, 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).

A report into the state of evaluation within US non-profit organisations (Innovation Network, 2013) found that the three most significant barriers to evaluation were:

1. lack of staff time
2. insufficient financial resources
3. limited staff expertise in evaluation.

The most important resource for evaluation departments is their personnel. Unfortunately, staff in evaluation departments are not always as well supported as they should be. '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). One agency report from 2007 noted that '[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).

[12]Riccardo Polastro, DARA, personal communication, May 2013.

Given such incentive structures or lack of them, it may be that evaluation departments will not perform at optimal levels. Evaluation staff should be trained to an appropriate degree, and have access to skills building as well as refresher training opportunities as forms of continuing education on evaluation.

Evaluator insight 10

DFID's initiative to increase support for evaluation

The recent UK Independent Commission for Aid Impact (ICAI) synthesis study of DFID's strategic evaluations found that limited staff skills in M&E was a key obstacle in evaluation uptake and use, and this was further highlighted by the UK's National Audit Office in its examination of DFID's performance management, which documented a lack of staff training in this area. Moreover, an evaluation of the DFID programme in the Caribbean concluded that M&E efforts had been hampered by a reduction in staff with skills in M&E (Drew, 2011: 59). Part of DFID's response, was to recognise evaluation as a profession in its own right, accrediting 133 staff members as evaluation specialists to one of four skill levels. There has also been an increase in opportunities for evaluation training, exchange and peer learning and the launch of an annual Evaluation Professional Development Conference.

Sources: Government Office for Science, 2011b; DFID, 2013; Jonathan Patrick, DFID, personal communication, May 2012.

Induction courses and external training events are often only one component of an employee's orientation to an evaluation-related job. Other key components concern less tangible elements such as: creating a common language, with shared conceptual references and ideas, and building a common understanding of the purposes and practices of EHA within that evaluation office. We found no EHA-specific research analysing the impact of evaluation training on evaluators' performance and quality of evaluation commissioned and carried out. Some observers noted the relatively few opportunities for training specifically in the evaluation of humanitarian action, and that, while beneficial, the impact of training occurs over time. Therefore, it is hard to find concrete examples proving direct benefit from attendance at training events[13].

[13]Margie Buchanan-Smith, Independent, personal communication, June 2013..

In addition to sending evaluation staff to evaluation training and refresher courses, staff rotation through the evaluation department has also been encouraged and is regularly applied in several organisations such as UNHCR and WFP. WFP for instance, has a policy of rotating staff every 4 years. In their Office for Evaluation, half of the staff are on rotation and half are hired as specialists from outside (WFP, 2008b: paragraph 48).

CAPACITY AREA 2

PURPOSE, DEMAND AND STRATEGY

Clarify the purpose of evaluation and articulate it in evaluation policy

Whether an evaluation can simultaneously meet both accountability and learning needs is a contentious issue. The evaluation literature tends to suggest that these two aims are in conflict, and many interviewees agreed.

Evaluator insight 11

Be aware of learning and accountability fault lines

A brainstorming session among experienced humanitarian evaluators at an evaluation workshop found that an overly strong *upward* accountability focus in evaluations was a factor in their poor utilisation. A strong accountability focus led participants to feel judged and less willing to engage during the evaluation process and afterwards with the findings and recommendations.

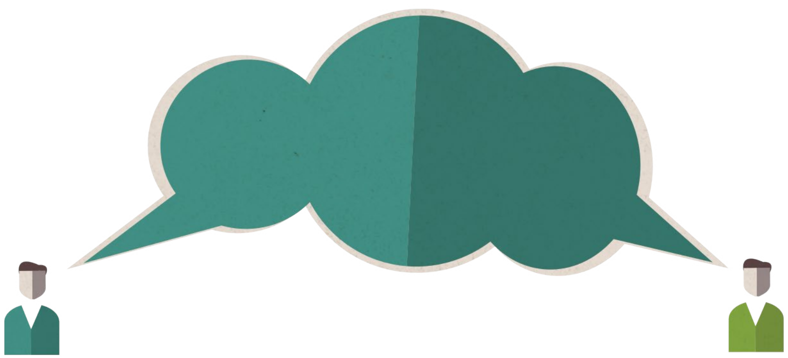
Source: ALNAP Evaluators peer learning event, Washington DC, 2011.

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, 1997).

Other observers see this accountability and learning fault line less starkly. While evaluations often emphasise one of these two purposes, they can, in practice, contribute to both. Indeed, some are adamant that an evaluation must address both aims, and that one can only truly learn by being held accountable. The 'truth' claim will vary by organisation and culture, and there is no single 'correct' approach here. For example, the UK Disasters Emergency Committee (DEC) is in the process of developing its 'policy and approach to learning'. As is laid out in this document, even where accountability is the prime objective of an evaluation or study it has commissioned or funded, 'this should not preclude opportunities for learning where they are apparent' (DEC, 2013: 1).

Some organisations do seem unclear about the purpose of evaluations, and about recognising that different approaches may be required for different evaluation aims. 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 difficult to achieve.

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.



> *Just as no single evaluation can serve everyone's interests and answer everyone's questions… no single evaluation can well serve multiple and conflicting purposes. Choices have to be made and priorities established. But instead, evaluators, those who commission and fund evaluations, and intended users persist in expecting evaluations to serve multiple and competing purposes equally well. Those unrealistic expectations, and the failure to set priorities, are often the place where evaluations start down the long and perilous road to incomplete success (Patton, 2010b: 151-163.)*

One way to mitigate the tension between the differing aims of evaluation is to separate 'accountability-oriented' evaluations from the more 'learning-oriented' evaluations, and not try to meet all agendas with one exercise.

> *If evaluations were seen less as a threat and more as a 'learning tool' for managers to assist them to deploy their resources more effectively, then, on the one hand, evaluators would be encouraged to write less cautious and more useful reports, while, on the other hand, managers would be likely*

> *to take more interest in engaging more deeply with the recommendations made and their follow-up. (Riddell, 2009: 13)*

Arguably, it might 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. Taking the argument further, one could argue that perhaps, *upward* accountability evaluations could be carried out by the audit and accountability department, whereas field-led learning and forward accountability-oriented evaluations could be carried out by a department of knowledge management and learning.

Different organisations have been adjusting task allocations and updating job descriptions in an attempt of striking a balance between the different evaluation, learning, accountability and audit functions. The UK DEC, for instance, after an initial experience with appointing an Accountability and Audit Manager and delegating to another staff the learning-related activities, has shifted towards a more joined-up approach. The organisation has appointed an Accountability and Learning Officer and made the Head of Programmes and Accountability responsible for both elements of programme reporting (evaluation as well as learning).

Evaluator insight 12

Who delivers on management and accountability responsibilities?

Most evaluation commissioners and contractors continue to conceive evaluation as a 'tool for accountability' with accountability understood in the classic sense of 'the obligation to account for a responsibility conferred'. This gets it all wrong. Accountability is an integral management responsibility, i.e. it is up to management to report on its performance, and when a third party does it, management is not only off the hook, it is not managing fully. Having a third party take on the accountability responsibility is questionable, all the more in so called results-based management contexts. And if individuals somewhere want assurance on performance or performance reports, then they should use performance audit, not evaluation.

Source: Ian Davies, XCeval evaluation listserv, January 2013.

Just as no single evaluation can serve everyone's interests and answer everyone's questions... no single evaluation can well serve multiple and conflicting purposes.

The evaluation need depends on the context. The 'right' approach also depends on the organisation's receptiveness and capacity for learning lessons. In doctoral research where experienced evaluators were asked to determine the best type of evaluation for different organisations. It was recommended that organisations less ready for learning invest more in carrying out process evaluations. Conversely, those more open to change should focus more on outcome-oriented evaluations (Allen, 2010).

Once there is clarity, this can be reflected in an organisation's evaluation policy, which can then be communicated throughout the organisation. ACF has recently developed a new evaluation policy which has brought together the three key ideas at the core of their vision for evaluation and learning (Guerrero, 2012; ACF, 2011b):

1. Evaluations are only valuable if they are used.
2. Evaluations must help track progress.
3. Evaluations must be an integral part of organizational learning.

Increase the demand for evaluation information

Ideally, the supply of evaluation should match the demand for evaluation products. There are several steps evaluators can take to increase the demand for evaluation information (Box 5). An example of an organisation that works to foster these cultures through demand for evaluation findings is the DEC (see evaluator insight 13).

Evaluator insight 13

DEC focus on learning from previous responses

Learning is at the heart of how the DEC works; it is one of the four pillars of its Accountability Framework (DECAF) and one of its priorities for how members work (DEC, 2013: 1). This latter component is nicely reflected in how members are asked from the very beginning of the response cycle to consider past lessons learned. This is to say that while planning their response members are asked to consider what learning from previous interventions will be applied. Though this is not perfect, as it may become a box-checking exercise or be passed on to the evaluation department, it aims to encourage reflection on past learning and, hence, increase overall utilisation.

Source: Annie Devonport, DEC, personal communication, June 2013.

Box 5

What factors influence policy-makers' demand for and use of evaluative evidence?

Openness – of the system to the entry of new ideas and democratic decision-making.
A champion – at a high political level.
Awareness – general understanding among policy-makers of the utility of M&E data/findings.
Utility – having a link between decision-making and the M&E system increases the perceived utility of information.
Incentives – for usage, e.g. including evaluation use in performance plans and publicly recognising those who demonstrate good evaluation use.
Donor influence – specifically, donor pressure on the host country to invest in evaluation.

Source: Levine and Chapoy, 2013: 5.

A recent study commissioned by 3ie looked at commitment to evaluation as a possible indicator to capture the readiness and willingness of aid agencies to demand evaluation knowledge, commission, conduct and use evaluation. The authors propose that a:

> *commitment to evaluation is demonstrated when agencies commission credible evaluations on a systematic basis, making the evaluations publicly available and designing programmes that are in line with evaluation recommendations.*
> *(Levine and Chapoy, 2013: 3-4)*

However, in addition to focusing on creating a receptive culture for evaluations, there are other practical steps to increase demand. Involving stakeholders (and identifying primary intended users as part of this process) and making sure that evaluations are timely and an integral part of the planning process are two such steps, as discussed below.

Strive for stakeholder involvement

Given finite resources, it is necessary to set priorities for stakeholder engagement, and to seize those opportunities that may prove more valuable[14]. After identifying intended users, as Patton (1997) explains in his writing on utilisation-focused evaluation, we need to identify also their explicit commitments to possible uses of evaluative knowledge[15]. Yet, intended users are more likely to progress findings if they understand and feel ownership of the evaluation process and findings. As confirmed during this study, one sure-fire way to create this ownership is to involve users actively in the process, thus building a relationship between them and the evaluation team or the evaluation itself. Patton (ibid.) suggests that evaluation should start with the generation by end users of questions that need to be answered.

[14]Bonbright, in Use of Impact Evaluation Results, suggests asking the following questions:

- For this evaluation, with these purposes, which stakeholders are the most important users?
- What do they have at stake in the evaluation?
- How powerful are they in relation to other stakeholder groups?

He goes on to suggest that provision needs to be made to support low-power groups who have a high stake in the evaluation (2012). Also see Buchanan-Smith and Cosgrave (2013) sections 2.5 to 2.7.

[15]For more information on the topic of stakeholder engagement in evaluation, also see Weiss (1983b). For more recent perspectives see Forss et al. (2006) and Oral History Project Team (2006).

Evaluator insight 14

Sida's evaluation planning cycle

The process starts with the evaluation unit having conversations with all operational units, to determine their knowledge needs, what they want to know and how evaluation could help. This generates a list of around 100 ideas for evaluations. From this, the evaluation department chooses 15 evaluations to 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 the evaluation process.

Source: Joakim Molander, Sida, in Hallam, 2011.

FAO systematically involves stakeholders from the beginning of thematic or country evaluations, using consultative groups including affected people, major donors and programme personnel. The group ensures that the right questions are asked, and has a chance to comment on draft products. This practice helps to break down the often supply-led approach to evaluations, particularly those done for compliance purposes (Box 6).

Box 6

Who makes up your evaluation audience?

Generating internal demand can be a challenge when evaluations are required from above. Researchers investigating the state of evaluation in the USA asked a sample of NGOs to identify the primary audience for their evaluations. The most frequent primary audience was funders, followed by the organisation's boards of directors and other organisational leadership. Other staff and clients were rarely mentioned as the primary audience. Organisations committed to evaluations only as a condition for accessing funding are unlikely to value or have 'ownership' of the resulting products.

Source: Reed and Morariu (2010).

Evaluator insight 15

DEC experience in building stakeholder ownership around evaluation work

The DEC aims to build ownership by involving its members at numerous steps during the evaluation process. First, the ToRs of its Response Reviews (previously known as DEC Real-Time Evaluations) are widely circulated for comment and review by members. Members are asked to share these with stakeholders to determine the 'why' of the evaluation, what they would like to know or gain from this exercise. Thereafter, once in the field, the consultant is required to hold two meetings with agency representatives and field teams, be these staff or partners. The first is to clarify the evaluation's purpose and desired results, while the second is to feed back findings.

Source: Annie Devonport, DEC, personal communication, June 2013.

Both DFID and CARE have seen positive results from even minor increases in user involvement. DFID found that interviewees' active involvement with the Evaluation Unit improved their understanding of the department and helped break down misconceptions about being 'audited' (Jones and Mendizabal, 2010). Similarly, a study on utilisation within CARE found that being included in the evaluation process, as an interviewee or participant in an After Action Review for example, heightened the individual's sense of ownership in the recommendations (Oliver, 2007: 2), made them more interested in findings and more likely to see these as 'a legitimate source of information for their practice' (Oliver, 2009: 36-37).

Evaluator insight 16

Making the most of country office involvement in evaluation

To maximise utility at country level, the UNICEF Evaluation Office sought to involve the country office (CO) at all stages of the evaluation through the formation of a local reference group. The evaluation manager went to Timor-Leste and involved the CO M&E officer in discussions around sampling strategy, survey design, and offered support with evaluation quality assurance to boost the overall quality of the evaluation.

This element of capacity development served as an incentive for the CO M&E staff to support the education programme actively, and provided them with a stake in its success.

The result was a high-quality, relevant evaluation that led to a strong country office-led management response and follow up process. As a consequence of this independent yet collaborative approach, the donor committed a significant second tranche of multi-year funding for the programme. This process also changed the culture in the country office, which became much more supportive of evaluation.

Source: UNICEF, 2010b

Patton argued in a recent edition of the *Canadian Journal of Program Evaluation* (2010b), that incomplete successes and shortcomings in evaluation work inevitably involve — at some level and to some extent — difficult and challenging relationships with stakeholders. Experienced evaluators suggest some strategies here: engage stakeholders sooner; and engage with them more clearly, building their capacity, managing relationships skilfully, and dealing with problems in a timely way before they escalate to crises.

Ensure evaluation processes are timely and integral to the decision-making cycle

An important influence on evaluation demand is timing of the evaluation itself and of the presentation of its findings. Many potential evaluation users complain that evaluations often arrive too late to be of any use in decision-making (Weiss, 1983b; 1990).

A report on DFID 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. A senior evaluator in MSF notes:

> *While 'timeliness' is such an important criterion in many of our evaluations, the process itself is often painfully affected by the 'speed of action' in MSF. This means that change of priorities*

> *and turnover of people can make an evaluation irrelevant right after it has been completed.* [16]

It is not unusual for evaluations to finish (or even start) only after the peak of an emergency is over. There are two difficulties here. First, significant analysis becomes more difficult as time passes, as the memories of key informants wane and information is lost. Second, long delays between programmes and evaluation conclusions means that there is then no opportunity to use the findings to change the programme concerned.

It is not unusual for evaluations to finish (or even start) only after the peak of an emergency is over.

In a publication on how to maximise learning from evaluation, WFP (2009) reported that staff frequently complained of information overload, saying they had insufficient time to read even a two-page newsletter. At the same time, a 16-page document produced by ALNAP on lessons for operational agencies on flood response[17] was widely read. The explanation for this, based on literature review and experience from agencies, appears to be the need for a good understanding of what is most useful to the targeted user, and finding the right combination of content, timing and format.

As the Head of Evaluation at UNHCR puts it:

> *Timing is a key factor in determining the impact of an evaluation. Part of the 'art' of an evaluator is to scan the horizon, to look forward and to anticipate those issues that will be (or should be) of most concern to senior management in the near future. This is more important than timing evaluations to coincide with the programming or funding cycle.*[18]

[16]Sabine Kampmueller, MSF, ALNAP Evaluation CoP contribution, October 2012.

[17]Alam, K. (2008) 'Flood disasters: learning from previous relief and recovery operations'. London: ALNAP/Provention consortium (www.alnap.org/pool/files/ALNAP-ProVention_flood_lessons.pdf).

[18]Jeff Crisp, UNHCR, ALNAP Evaluation CoP contribution, July 2012

Concerns such as these have led to the recent growth in the number of real-time evaluations (RTEs) by a range of humanitarian organisations. An 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.

A real-time evaluation of the UNICEF response to the 2012 nutrition crisis in the Sahel was launched just three months after the declaration of the crisis, to provide rapid and early learning to improve future response and to help mitigate future crises. This RTE enabled UNICEF to establish a more coherent framework to guide its response, including concrete recommendations on how to reach more children by extending the number of centres providing malnutrition treatment, using mobile clinics, promoting community-based case management of severe acute malnutrition, increasing capacity in remote areas, and better integrating service provision. The report's results have been widely used in the region, with country offices developing detailed management responses immediately after the exercise, and the regional office implementing the recommendations (UNICEF, 2013).

Evaluator insight 17

WFP's experience with integrating evaluation into policy development

In 2011, WFP carried out an evaluation of its school-feeding policy. This was designed to fit decision-making processes. The new school-feeding policy was adopted in 2009 and a condition of board approval of the new policy was that it should be reviewed by end 2011. This allowed for forward planning. Five individual country impact studies were carried out fed into the policy evaluation, which enabled the policy evaluation to be built on solid foundations, with a clear timetable to work towards. Demand for the evaluation was also pre-determined from the outset, to fit into the board review of the policy. WFP now has a policy that new policies be evaluated within four to six years of implementation, so future demand for evaluation products around the policy issue in question is guaranteed. The same approach was taken ahead of the review of WFP's strategic plan. In 2010, the WFP Office of Evaluation planned four strategic evaluations which all looked at different dimensions of the shift from food aid to food assistance (the core component of WFP's Strategic Plan 2008-2013). This required a considerable degree of pre-planning but the use of the end products was considerable.

Source: Sally Burrows, WFP, personal communication, March 2013.

CARE found that the evaluation and After Action Review for tropical storm Jeanne in Haiti was a model of how evaluation can effectively inform planning and preparedness due partly to the consideration given to its timing. An initial review within six months of the storm allowed for the participation of a good cross-section of staff, despite the fact that some had already departed, while the distribution of the final evaluation report in March allowed for its use in CARE Haiti's annual planning event in April that year. The report identified resource gaps, such as storage and distribution points for potable water that the planning session was able to address for the following fiscal year. Finally, the report was useful in scenario-building and subsequent contingency planning (Oliver, 2007: 14).

Develop a strategic approach to selecting what should be evaluated

There is a 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 (Box 7).

Box 7

Is evaluation always warranted? Wisdom from Carol Weiss

If evaluation results are not going to have an influence on how people think about the programme, it is 'probably an exercise in futility'. In one of her classic works, Carol Weiss discusses four kinds of circumstances in which evaluation may not be a worthwhile exercise.

1. When the programme has few routines and little stability: if it is not clear what the programme is and how it is meant to operate, it would not be clear what the evaluation means.
2. When people involved in the programme cannot agree on what the programme is trying to achieve.
3. When the sponsor of the evaluation or programme manager sets stringent limits to what the evaluation can study, putting off limits some important issues.
4. When there is not enough money or staff sufficiently qualified to conduct the evaluation.

More positively, Weiss identifies other circumstances where evaluation can be a worthwhile undertaking. These include:

- mid-course corrections
- continuing, expanding, institutionalising, cutting, ending, or abandoning a programme
- testing a new programme idea
- choosing the best of several alternatives
- deciding whether to continue funding
- understanding programme-intervention logic
- accountability for funding received and resources used.

Source: Weiss, 1998: 24-28.

Some evaluation departments plan evaluations on the basis of ensuring that all major programmes are evaluated every few years. The guiding principle of such an approach 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. Similarly, allocating a fixed percentage of programme costs to evaluation budgets may not be the most cost-effective way of using scarce resources.

It may be better to take a strategic approach when selecting what should be evaluated (Box 8). This may entail proportionately allocating more resources to more complex programmes, or to those perceived as more contentious, or those where evidence of effectiveness is less readily available. This may also lead to allocating more evaluation resources to smaller programmes, rather than to large ones that have been running for years. Experience suggests there is value in trying to be more explicit and intentional when selecting intended users, possible uses and timing to undertake evaluative work.

WFP's 2011 Annual Evaluation Report noted an increased focus on higher-level evaluations: global evaluations (strategic and policy), country-portfolio evaluations, and a series of impact evaluations. It carried out only one evaluation of a single operation in 2011. The report stated that, given limited evaluation resources, this approach promises greater value-added to WFP by providing evidence to inform strategic-level decisions regarding policies, country strategies or corporate strategies, while impact evaluations provide more in-depth assessment of outcomes, impact and unintended effects than do single-operation evaluations.

Box 8

Sharpening the evaluation focus in UNICEF

In 2012, UNICEF was the agency with the most cluster-related areas of responsibility, and so undertook a 'systematic, in-depth and independent assessment of its performance in managing its Cluster Lead Agency Role (CLARE) in Humanitarian Action. The first of its kind, this evaluation was a response to a great variety of internal and external pressures and aimed to generate evidence of results – or lack of them – to aid management in improving UNICEF's Cluster leadership and coordination function (UNICEF, 2012c: 2-4).

An important component of this evaluation was a series of country case studies. In order to reduce the margin for bias and subjectivity in the country selection, the humanitarian evaluation team at UNICEF used a 'cluster analysis' – a statistical technique to select the country case studies more systematically and objectively. Country offices were grouped according to a comprehensive list of variables such as: emergency profile (type and level); stages of Cluster arrangement implementation (i.e. early activation, implementation, phasing out); number of Clusters in place; number of Cluster members at national and sub-national levels; presence of OCHA office; presence of a UN peacekeeping mission; capacity and engagement of a national disaster management authority; and availability of emergency funds.

The evaluation office then identified eight groupings of country offices with homogeneous characteristics, and proceeded to select one case study country from each (UNICEF, 2012b: 1, 3). Moreover, the Evaluation Office attempted to ensure maximum UNICEF regional representation and avoid evaluation fatigue in country offices frequently evaluated. This systematic approach led to a more thematically relevant and representative set of country case studies.

Source: Erica Mattellone, UNICEF, personal communication, March 2013.

The DAC peer review of DFID's evaluation function concluded that its M&E requirements were too complex; some reviews were time-consuming and became rapidly out-dated as external circumstances changed (Riddell, 2009). This review also found that the volume of individual DFID evaluations was in danger of outstripping the capacity of most stakeholders to assimilate lessons. A World Bank report came to similar conclusions:

> *the utility and influence of many methodologically sound evaluations has been limited because they were looked upon as one-off evaluations and did not form part of a systematic strategy for selecting evaluations that addressed priority policy issues or that were linked into national budget and strategic planning.*
> *(World Bank, 2009: 70).*

CAPACITY AREA 3

EVALUATION PROCESSES AND SYSTEMS

Strengthen pre-evaluation processes

Strengthening pre-evaluation processes is vital. A World Bank publication which calls this 'front-end analysis' describes the following.

> *Many evaluators are impatient to get the evaluation planning finished and therefore rush into data collection. They try to do exploratory work at the same time as data collection. But completing a good front-end analysis is critical to learning about an intervention. It can save time and money on the evaluation, ensure the evaluation meets client needs, and sustain or build relationships, not only with the client but also with key stakeholders. Most important, a good front-end analysis can ensure that the evaluation is addressing the right questions to get information that is needed rather than collecting data that may never be used.*
> *(Morras Imas and Rist, 2009: 142)*

In addition to noting the importance of user engagement and timing, the World Bank study recommends considering the following questions (Morras Imas and Rist, 2009).

- How much time is available to complete the evaluation?
- What resources are available for the evaluation?
- Does social science theory have relevance for the evaluation?
- What have evaluations of similar programmes found? What issues did they raise?
- What is the theory of change behind the project, programme or policy?
- What existing data can be used for this evaluation?

One of the most important pre-evaluation requirements is that the programme being evaluated is 'evaluable'. This requires programme managers to ensure that there is a performance framework in place before a programme is implemented. Such a framework should, in the words of Riddell (2009), provide clarity about what success is expected to 'look like', identify and specify the quantitative and/or qualitative evidence that will be monitored and used to judge performance, including value for money, and address the issue of attribution. The same goes for policy and strategy development.

Looking at DFID for instance, Riddell (2009) found that in the past, one of the weaknesses of DFID's approach was the lack of consideration of future review or evaluation built in when key policies and strategies were developed and (eventually) launched. The result was that all evaluations tended to be ex-post in nature. This has now changed, and DFID adopts a prospective approach to evaluation, with all new interventions requiring a decision on whether an independent evaluation is needed from the outset. There is a strong expectation that all interventions over £5m, or those that are new or innovative, will have a strong evaluation built in from the start and incorporated in the business case. The aim of these changes has been to ensure both quality assurance and the 'evaluability' of DFID-funded programmes before their implementation (Government Office for Science, 2011a: 14; York, 2012: 19).

In 2012, UNICEF received funding from DFID to boost its capacity for humanitarian action. As well as being expected to achieve results in four specific capacity-related areas, UNICEF was also required to be able to demonstrate this achievement (UNICEF, 2013: 12-13). The Evaluation Office commissioned an 'evaluability' assessment asking the following questions.

- Was the programme logic clear?
- Was it clearly communicated to stakeholders and understood and accepted by them?
- Were planned activities strategically sound and were resources sufficient?
- Were management arrangements optimal?
- Would the data being collected demonstrate whether results were being achieved?

This assessment helped management strengthen the linkages between programme activities and targeted results and led to a more strategic process of allocating programme funds. The exercise demonstrated to UNICEF how evaluation and management can work together to address issues before they become problems (ibid.)

Improve the quality of evaluation

Delivering high-quality evaluation products may stimulate an organisation's 'appetite' for maintaining higher quality standards in evaluative work.

> *Quality is central to ensuring the credibility and utility of evaluations. It is manifest in the accurate and appropriate use of evaluation criteria, the presentation of evidence and professional analysis, the coherence of conclusions*

with evaluation findings and how realistic the evaluation recommendations are.
(WFP, 2008: 9)

ALNAP's new pilot guide Evaluating Humanitarian Action goes into much more depth on these issues (Cosgrave and Buchanan-Smith, 2013). This section of the present paper does not seek to duplicate this work, but highlights some overarching themes that emerged from the interviews and research.

Limit the focus of evaluation

- 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, the scope of the evaluations 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.
- An ICAI review of the quality of DFID evaluations found that some evaluations lacked focus and tried to cover too many issues without prioritisation. This resulted in evaluation reports that were very long, limiting the ability to use them (Drew, 2011).

Involve beneficiaries

As discussed in Section 2 above, engaging stakeholders and more specifically users of evaluations, improves utilisation. Those affected by crisis, however, should arguably be at the centre of not just the evaluation process, but of humanitarian action (IASC, 2012). Although there is still a long way to go in achieving this, humanitarian organisations are working to improve the current situation (ALNAP et al., 2012). It is now slowly becoming the norm for affected populations 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[19]. Further to engaging with affected populations, some evaluation exercises have emphasised involving host-government counterparts (Box 9).

[19]See for instance Catley et al., 2008; ALNAP and Groupe URD, 2009; Panka et al., 2011; Anderson et al., 2012.

Box 9

Making the most of in-country evaluation stakeholder workshops

This was the first IASC RTE to use workshops as the main feedback mechanism, allowing people on the ground to participate actively in the framing of recommendations, and moving this process away from an HQ-imposed exercise (www.youtube.com/watch?v=qW-XR14k4NQ). It seems significant that most of the Pakistan Inter-Agency RTE recommendations have subsequently been adopted. Participants found this process helpful in thinking through what they needed to do, boosting real-time learning and peer-to-peer accountability in the field. Government involvement in this process also contributed to assigning priorities to different follow-up actions.

As part of the Inter-Agency Standing Committee (IASC) real-time evaluation of the response to the floods in Pakistan (IASC, 2011a), four provincial workshops were held in Karachi, Punjab and Islamabad (for KPK and Baluchistan provinces). Some 20–40 participants from the government, international and local NGOs, UN agencies and the Red Cross Movement participated in each workshop. There was then a national workshop involving donors as well as some representatives of organisations that had participated in the provincial workshops. Recommendations were taken from one workshop to another, and were improved on each occasion. As a result, there was a building of consensus, refining of recommendations and clarity about priorities, timing and who should do what and at what level (provincial, national or global).

Each workshop took place over half a day. The total workshop process took up a full week of the evaluators' time, but the process was considered highly useful and a key reason for the consensus gained and the implementation of recommendations. In addition, recommendations could be implemented immediately without the need to wait for the final evaluation report to be issued. Once endorsed and prioritised in the field, the conclusions and recommendations were presented in Geneva and New York, enabling rapid dissemination and policy take-up at headquarters level. Together with other key inter-agency evaluations, this RTE fed into the IASC Transformative Agenda (IASC, 2012).

Source: Riccardo Polastro, DARA, ALNAP Evaluation CoP contribution, July 2012.

Quality assurance

A number of agencies seek to improve evaluation quality and standardise approaches through quality assurance systems and/or external review processes. WFP uses an approach called Evaluation Quality Assurance System (EQAS), which focuses on immediate evaluation outputs and gives evaluation teams clear guidance on the standards expected. It also gives evaluation staff clear guidance on timing and what standards they should be applying. EQAS does not currently look at wider management processes around evaluation, but a review is underway and consideration being given to expanding the quality assurance process into evaluation management systems, such as selecting teams, carrying out post-evaluation processes, and ensuring that learning opportunities are maximised.

Although evaluation managers have taken some time to embrace it fully, EQAS is being recognised as an asset to the organisation. Evaluation teams appear to appreciate the clarity of expectations it brings from the outset of a piece of evaluative work, though some have struggled with its perceived lack of flexibility.

Other UN agencies have also expressed interest in adopting a similar approach (WFP, 2009)[20].

Standardized EQAS requirements have improved the quality of evaluation reports, and the collaborative process used to develop the materials increased understanding and application of these standards. In 2011, the Office of Evaluation expanded its use of external reviewers for those evaluations with especially high levels of diverse stakeholder interest. These review panels are separate from the independent consultants who conduct evaluations, and provide an additional dimension for the quality assurance of methodology and/or content (WFP, 2011).

[20] Sally Burrows, WFP, personal communication, March 2013.

Evaluator insight 18

UNICEF's evaluation quality assurance and oversight system

UNICEF has introduced a Global Evaluation Reports Oversight System[20] (GEROS) which provides senior managers with a clear and short independent assessment of the quality and usefulness of individual evaluation reports, seeks to strengthen evaluation capacity by providing real-time feedback on how to improve future evaluations, and contributes to corporate knowledge management and organisational learning, by identifying evaluation reports of a satisfactory quality to be used in meta-analysis. The assessment of final evaluation reports is outsourced to an external, independent company. An evaluation report is assessed as satisfactory when it is a credible, evidence-based report that addresses the evaluation's purpose and objectives and can, therefore, be used with confidence.

One of the criteria used for judging whether an evaluation report is satisfactory is that the purpose of the evaluation is clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information was needed and how the information was to be used. UNICEF evaluation personnel have noted that where GEROS deems an evaluation to be of high quality, managers have increased confidence to act on the findings and that their commitment to evaluations and implementation of their recommendations is strengthened.

Source: Rob McCouch, UNICEF, personal communication, March 2012.

[20] UNICEF, 2010c; see also: www.unicef.org/evaluation/index_60830.html

Engage in peer-review of the evaluation function

In recent years, and particularly among UN agencies, there has been an increase in conducting peer reviews of the evaluation function – in which colleagues from similar organisations come and assess the setup and overall performance of the evaluation office. At NGO level, however, there has so far been little experience with peer reviewing the evaluation function of affiliates or sister organisations. Yet, this is taking place in other areas of interest to NGOs.[22] UN peer review has involved looking at three core issues: credibility, independence and utility (UNEG, 2005), as detailed in Box 10. The purpose of these peer reviews has been to build greater knowledge, confidence and use of evaluation systems by management, governing bodies and others, provide a suitable way of 'evaluating the evaluators' and to share good practice, experience and mutual learning (DAC/UNEG, 2007; UNEG, 2011).

Box 10

The evaluation principles trinity

In 2008, the UNEG framework for peer reviews outlined three principles to be central to evaluation: credibility, independence and utility.

Credibility is concerned with the competence of the evaluators, transparency of the evaluation process, and impartiality of evaluators and process.

Independence means that the evaluation is free from political or organisational pressures or personal preferences that would bias its conduct, findings, conclusions or recommendations. It implies that evaluations are carried out or managed by entities and individuals who are free of the control of those responsible for the design and implementation of the subject of evaluation.

Utility means that evaluations aim to and do affect decision-making. Evaluations must be relevant and useful and presented clearly and concisely. Evaluations are valuable to the extent that they serve the information and decision-making needs of intended users, including answering the questions raised about the evaluation by the people who commissioned it.

Source: Heider, 2011 (based on UNEG, 2005; UNEG, 2008; WFP, 2008).

[22]For example: performance (NGO and Humanitarian Reform Project, 2010), programmatic (IFRC, 2011a; Cosgrave et al., 2012; CBHA, 2012) and thematic areas, such as accountability (SCHR, 2010).

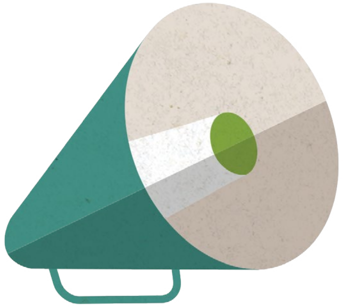
The first UN peer review was undertaken in 2005, and this practice has since been a regular feature of UNEG's work. The process was conceived initially to deal with the fact that bilateral donors were not seen to be using the reports and results of evaluation work carried out by multilateral agencies (UNEG, 2013). The aim was to establish the credibility of evaluation reports coming from the organisation itself and thus to decrease the need for external multidonor evaluations of an agency or its evaluation office. The main advantage has been peer learning and peer exchange, and, in particular, the opportunity for evaluators from different agencies to learn about each other's realities and practices. Raising the profile and highlighting the value of the evaluation function in a given organisation – especially in the eyes of policy and strategy teams, as well as governing bodies – has also be seen an important benefit (ibid. 13).

Disseminate findings effectively

Evaluation products, and evaluative information in general, are of potential interests to very diverse audiences; this may include: programme participants and affected communities, programme managers, funders, advisors, policy makers, government officials, researchers and academics, international and local media, etc[23]. Ideally, each audience should receive a communication product that presents and 'packages' evaluative information tailored to their particular learning and information needs (Hallam, 1998; Proudlock and Ramalingam, 2009: 13; Morra Imas and Rist, 2009: Chapter 13; Cosgrave and Buchanan-Smith, 2013: Section 7).

Along these lines, organisations may want to consider having clear policies on customising information products as well as on communicating and disseminating evaluation deliverables. For example, as is laid out in the IFRC 'Framework of evaluation', ToRs are to include an initial dissemination list 'to ensure the evaluation report or summary reaches its intended audience. …The dissemination of the evaluation report may take a variety of forms that are appropriate to the specific audience' (IFRC, 2011a: 15). This includes an internal and external version of the report. Strategic dissemination of findings is crucial to making evaluations more effective.

[23]See Buchanan-Smith and Cosgrave (2013), in particular Sections 2.5 for an explanation of stakeholder mapping and 2.6 for a discussion on the importance of focusing on what primary stakeholders need to know.



Strategic dissemination of findings is crucial 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, how it is presented, and contributes to ensure it meeting the information needs of decision-makers.

Jones and Mendizabal (2010) have 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 inter-relations 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'. (See Box 11)

Box 11

Seven new ways to communicate evaluation findings

1. Summary sheets or research/policy briefs. A shorter document is much more likely be read than the full report.
2. Findings tables. There a risk of dumbing down, but presenting the raw findings can communicate your messages very strongly.
3. Scorecards or dashboards for real-time monitoring.
4. Interactive reports, interactive web-pages or web apps (e.g. http://www.ushahidi.com/).
5. Photostories, comic strips, info graphic, illustrations.
6. Blogs can be used in the process of evaluations as well as for discussing use.
7. Multimedia video report.

Source: expanded from O'Neil, 2012

Evaluator insight 19

High impact of end-of-mission reports

One evaluator notes that the most useful part of the evaluation process is the end-of-mission report. It is often written quickly and is short, without the extensive 'padding' found in finished evaluation reports. It is also available to those being evaluated very quickly after the event. Most importantly, because it is done so quickly after the event, and because it is not to be made public, is often full of passion, and able to generate a real debate within and response from the country teams involved. Circulation of the final report never produces as much debate and learning.

Source: Luz Gómez Saavedra, Oxfam Intermón, personal communication, March 2013.

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.
- USAID has commissioned a film about the evaluation process (Box 11).
- ALNAP itself produces a series of thematic lessons papers, based on distillations of evaluations. They are well received: the lessons paper on response to earthquakes was downloaded over 3,500 times within days of the January 2010 earthquake in Haiti.
- A fun way to present findings is to use the Ignite approach[24] in which all presentations are a maximum of five minutes long, with a new slide being automatically screened every 15 seconds. A surprising amount can be said in five minutes!
- Some evaluators are now blogging about their work, and there is interest in greater use of Facebook and Twitter to convey evaluation messages. (However, low bandwidth and internet access are critical issues for some country offices. Even where these are not a problem, information overload can be equally disabling.)

Box 12

USAID experimenting with film-making and learning from evaluation field work

USAID's Office of Learning, Evaluation and Research has recently commissioned a videographer, Paul Barese, to accompany an evaluation team in the field. The filmmaker followed a project evaluation team for over three weeks, to document most phases of the evaluation. One objective of this has been to generate learning about the evaluation process, convey some of the challenges in the field (particularly in conflict-affected areas), and allow those managing commissioning, designing and undertaking evaluations to better understand the complexity and value of evaluation.

The video is available at: www.usaidlearninglab.org/library/evaluating-growth-equity-mindanao-3-program

Source: Paul Barese in AEA 365 www.aea365.org/blog/?p=8768 Linda Morra Imas, IDEAS Evaluation listserv exchange, April 2013

[24]See: www.igniteshow.com/browse/popular

Box 13

Choosing evaluation dissemination strategies

Factors to consider when developing dissemination strategies:

- **Audiences** – Who will be receiving this? What are their special needs?
- **Rationale** – Why are you doing this? What do you hope to accomplish?
- **Content** – What type of information will it contain?
- **Purpose/use** – Why is this necessary? What could this be used for?
- **Timing** – When will this be completed? When will it be distributed?
- **Development** – What else needs to be done before this gets done?
- **Special issues/constraints** – What else is unique about this – must it be accompanied by other information? Are there certain requirements for accessing or understanding it?
- **Distribution** – In what ways will it be distributed? Will different people receive it differently?
- **Special concerns** – What are the limitations of this format?

These factors were used to analyse six dissemination strategies: full evaluation report, executive summary, PowerPoint presentation, fact sheet, statistics sheet and an online tool. Different dissemination outputs were found to have different advantages and disadvantages. The full report, for example, was read by very few people, yet was used as a resource in helping to resolve contentious issues. Other products, such as the three-page factsheets, were highly valued by fundraisers as a way of showing what had been achieved, yet were not used by the implementers at all. It was found that the more diverse the audience, the more individualised the dissemination format needs to be. Oral presentations were found to be powerful, especially when discussion of the results was encouraged.

Source: Lawrenz et al., 2007

Even where the traditional evaluation report remains, there are things that can be done to improve its chance of being used. A study by CARE suggested that, as well as shorter, more pointed evaluation reports, a ‘cover sheet’ for evaluation reports should be completed. This would categorise lessons-learned into specific areas (such as human resources, external relations or procurement) to facilitate the use of the findings by individuals responsible for these areas (Oliver, 2007: 18).

Interestingly, the same study also found that reports were not systematically shared and that very few of their interviewees seemed to know where to look if they wanted to locate a repository of reports. Clearly, however beautifully the report is presented, its use is limited if people cannot find it. A standardised format was also considered helpful for those wanting to skim the report rapidly (and would help evaluators in ensuring that their outputs were in line with CARE's expectations). Fields on the format could be linked to a searchable database to allow easy access to lessons learned in a concise format, either from individual evaluations or in the form of a synthesis (e.g. a summary of recommendations relating to human resources over the past two years).

Categorise lessons-learned into specific areas to facilitate the use of the findings by individuals responsible for these areas.

Strengthen follow-up and post-evaluation processes, including linking evaluation to wider knowledge management

Evaluation is just one of the tools with which an organisation can promote learning and change, as well as capturing new practice, knowledge and institutional memory. A recent paper from UNICEF on their humanitarian evaluation practice (UNICEF, 2013) finds that a dedicated knowledge management function is necessary to help staff learn and act on lessons from evaluations and other knowledge sources. UNICEF's emergency operation department (EMOPS) has attempted to integrate evaluation into policy initiatives for strengthening learning and accountability in emergencies by, for example, ensuring that evaluation concerns are incorporated into standard operating procedures, and by including the evaluation perspective in working groups on accountability and knowledge management. UNICEF recognises that the Evaluation Office needs the capacity to provide strategic guidance beyond management of requested evaluations. It also requires dedicated knowledge management capacity.

As Sandison reports in her ALNAP study on evaluation utilisation, 'evaluation planners and evaluators could do more to recognise the relationship between an evaluation and other learning processes' (2006: 135). While not attempting to cover the vast literature on learning and knowledge management (see for instance Hovland, 2003; Ramalingam, 2005), the discussion below touches on some of issues affecting the strength of the links between evaluation, knowledge management, follow-up and, ultimately, utilisation of evaluative data.

UNEG research and studies on evaluation follow-up have discussed how transparent evaluation management response and follow-up processes increase the implementation rate of recommendations (UNEG, 2010). The visualisation in Box 14 illustrates the linkages UNEG sees between evaluation processes and evaluation follow-up, as discussed in the following paragraph.

Box 14

Linking evaluation and follow-up processes

The figure below visualises research findings from recent work commissioned by UNEG (2010) to identify which preconditions are needed to support an effective management response and follow-up to evaluation, namely: high quality evaluation planning and implementation of evaluation. Throughout this study, a number of insights and features relating to high quality evaluation processes have already been introduced. In this context, 'high quality' should be understood as an attribute of evaluation process that are conducive to and support effective evaluation uptake and use[25] .

[25] A more comprehensive and humanitarian evaluation specific discussion on these features and evaluation process attributes is beyond the scope of this study, but is available elsewhere in the literature: see for instance Rey and Urgoiti (2005); ECB (2007); Groupe URD (2009); IFRC (2011b); MSF (2012); and Cosgrave and Buchanan-Smith (2013).

Preconditions for effective evaluation follow-up and management response

Good Evaluation Planning

Identification of key stakeholders, definition of evaluation focus, TOR preparation, evaluation team selection, logisitical arrangements for evaluation missions

Quality Evaluation Implementation

Briefings/inception events, evaluation field and desk work, report preparation, process for stakeholder comments and quality control of draft report (focus on quality and relevance of findings, lessons and recommendations)

Management Response to Evaluation

Management whose operations were evaluated provide a response, government and/or other partners may also respond to the evaluation

Disclosure and dissemeniation of Evaluation Report

Disclosure and publication (electronic a nd/or printed) of the evaluation, including management response; evaluation summaries or other knowledge sharing/ learning products

Follow-up to Evaluation

Formal and informal processes to promote, and verify, that evaluation based learning takes place within the organisation and among partners; management reports on status of implementation of recommendations

Source: UNEG, 2010:3

Ensure there is a management response to evaluations

Management response and follow-up to evaluations are essential for improving their impact (UNEG 2010). However, it is common parlance to refer to evaluations 'gathering dust on bookshelves', reports going unread, and with little formal or informal follow-up on findings and recommendations. When the same recommendations go disregarded or overlooked time after time, be it because of poor formulation, lack of actionable focus, timing issues etc. a perception may develop that evaluations are mainly done for symbolic or appearance's sake[26].

It has become common for agencies to have more formal system of management and response to evaluation aiming to reduce the chance that an evaluation process ends with the mere submission 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 (Hallam, 2011).

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 actions planned to address them. 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 include the opportunity to revisit evaluation findings, and to have a dialogue about the management response to them (Hallam, 2011).

Since 2011, the WFP Secretariat has organized an informal round-table consultation prior to each Board session, enabling more detailed discussion of the evaluation reports and Management Responses, to be presented formally at the session. This has

[26]On these points see for instance Weiss 1998; Patton 1997; McNulty 2012.

enriched the interaction between the Board and management concerning issues raised in the evaluation reports, and also enabled shorter, more focused discussion during the formal Board sessions (WFP, 2011).

The Independent Commission for Aid Impact includes a section in its annual report in which it presents the recommendations of each of its reports, whether these were accepted or rejected by DFID, a progress report on implementation of the recommendations and also DFID's response to the ICAI recommendations. The DEC on the other hand asks its members to share evaluation summaries and management response to evaluations work to be included on ALNAP's Evaluation Library. The aim is to create a culture of sharing learning with the wider humanitarian sector.

Evaluator insight 20

DEC post-mission workshops

To further encourage the utilisation of lessons learned from evaluations, the DEC holds post-mission workshops in the UK. These may be to share findings in general or they may focus on specific themes. In 2012 for instance, DEC co-facilitated a workshop to encourage disability and age inclusion in humanitarian responses. Interestingly, participants were asked to create a personal action plan, which will be followed-up after six months..

Source: Annie Devonport, DEC, personal communication, June 2013.

Ensure access to and searchability of evaluative information

Scanning, filtering and having the possibility of expeditiously accessing relevant information about evaluation activities current and planned as well as information about insights, findings, conclusions recommendations generated from evaluation work is surely an attractive prospect for evaluators.

Facilitating collection of and access to such evaluative information has the potential to improve the chances of evaluation resources being spent more effectively, and to reduce duplication in evaluation work. It could also improve the chances of evaluation findings and recommendations being looked at more systematically, referenced and taken up to inform decision-making. Box 15 presents three examples of recent initiatives to address some of these issues.

Box 15

Examples of increasing access information on evaluation

Tearfund has recently developed a Google Docs-based tool to share across the organisation details of planned and completed evaluations. Before an evaluation starts, staff charged with its management or supervision will be asked to supply details of: sectors covered, evaluation trigger (donor driven, collection of lessons learned, etc.), type of evaluation (mid-term, final or RTE), composition of the evaluation team (internal or external), desk- or field-based, approximate timeline and alignment to corporate outcomes.

Upon completion of the evaluation, components such as the management response plan, the contribution to corporate outcomes and key lessons will be added. The expectation and main objective is for this tool to: a) provide a snapshot of all evaluations planned and completed by Tearfund; and b) to facilitate knowledge sharing and communication across those engaging in evaluation activities and program staff interested in accessing evaluation findings and recommendations. Ownership of the tool rests with the M&E office at headquarter's level[27].

The recently launched **Humanitarian Genome** prototype aims to infuse innovation into the search, access and extraction of evaluation information. The project involves the development of a web-based search engine to sort and process individual humanitarian 'evaluation insights'. Evaluation reports have been digitised, and specific quotes (referred to as insights) extracted and then coded and scored. The score aims to determine the 'usefulness' of an insight and its ranking in the search findings, in a process similar to the search ranking function of Google or other search engines. It is noteworthy that the Humanitarian Genome allows users to search for information relating not only to evaluation content, but also to processes and methods[28].

When it comes to evaluation databases, one challenge is ensuring not only searchability but also completeness of the evaluative materials and reference entries. Established at the end of the 1990s, ALNAP's **Evaluation Library** offers to date the most complete collection of humanitarian evaluative materials – evaluation reports as well as evaluation methods and guidance material, and selected items on evaluation research. The aim of completeness and reliability of search results, however, relies on more than Network members submitting their evaluation reports. It also requires continuous efforts to seek and mine evaluation data from other sites in order to be as comprehensive as possible. The coding or tagging system needs to be revised frequently with the users' needs in mind, to optimise searchability. This includes providing intuitive search options and the ability to filter search results[29].

[27]Catriona Dejean, Tearfund, personal communication, July 2013

[28]Results Matter, 2013; see more detail at: www.humanitarianinnovation.org/projects/large-grants groningen

[29]More information available at: www.alnap.org/resources.

Conduct meta-evaluations, evaluation syntheses, meta-analysis and reviews of recommendations

Most potential users of evaluation results want to know more than just what one single evaluation or study found. 'They want to know the weight of the evidence' (Weiss, 1998: 317). Dozens of studies, evaluations and reviews may cover the same issue or theme. Looking at these in their totality, through meta-analysis and evaluation synthesis, can yield far richer evidence and findings. For example, a World Bank study on using impact evaluation for policy-making found that:

> *While individual evaluations may have made a useful contribution, the cases illustrate that the effects and benefits are often cumulative, and utilization and government buy-in tend to increase where there is a sequence of evaluations. In several cases, the first evaluation was methodologically weak… but when the findings were found useful by the national counterparts, this generated demand for subsequent and more rigorous evaluations. (World Bank, 2009: 70)*

The demand for evaluation synthesis, thematic reviews, and meta-analysis in the humanitarian sector appears to be on the rise. Every two years, the ALNAP Secretariat commissions the authors of the State of the Humanitarian System report to conduct a meta-analysis of all the humanitarian evaluation reports collected and stored – in that two-year period – in the ALNAP's Evaluation Library (see Box 15). The meta-analysis is the starting point of the subsequent investigation and analysis that look at, and track over time, humanitarian system-wide trends and performance against the OECD-DAC criteria.

Other agencies have been exploring consolidated approaches to the analysis of evaluations.

- Oxfam GB has started periodically conducting meta-evaluations of its work using set criteria (Oxfam GB, 2012; n.d.).
- CARE has carried out a meta-review of evaluations and after-action reviews from 15 emergency responses, and drawn together key learning from this (Oliver, 2007: Annex 1A).
- Every year, NORAD produces a synthesis report of lessons from evaluations[30].
- An ODI paper on learning lessons from research and evaluations within DFID (Jones and Mendizabal, 2010) found glowing references from interviewees to research and evaluation that offered synthesis and comparison of work on particular themes.
- In 2012, ACF established an Emergency Response Learning Project (Box 16).
- The RC/RC movement held an organisation-wide 'Learning Conference' on their Haiti Operation in 2012[31]. The 2013 event will include a meta-analysis of all programme and technical-level evaluations and reviews conducted by the IFRC, the Haitian Red Cross Society or any of the other National Societies participating in the Haiti Earthquake Operation. The aim is to use evaluation synthesis to capture good practice and technical improvements that can inform future updates of Standard Operating Procedures for the RC/RC movement (IFRC, 2013).
- CIDA has started to publish an annual report of lessons it has learnt from evaluations in the preceding year. The report groups together key findings, in a user-friendly and easy-to-consult fashion, and finishes with one page summaries of all the evaluations reviewed (CIDA, 2012).

[30] See: www.norad.no/en/tools-publications/publications/search?search=&region=&theme=&year=&type=annualreports&serie=&publisher

[31] The Haiti 2012 Learning Conference report is available on the IFRC Learning wikispace: http://ifrc-haiti-learning-conference.wikispaces.com/home

Box 16

ACF Emergency Response Learning Project

ACF International established an Emergency Response Learning Project in 2012 to assist the organisation in the development of more timely, appropriate and effective responses to the humanitarian emergencies of the 21st century. ACF used external evaluations of two recent emergency responses (Asian Tsunami in 2004 and Haiti Earthquake in 2010) to identify key themes around which emergency responses are built, and then to identify best practices and principles that had either been proven to deliver a more effective response, or that featured as recommendations in the evaluations. The aim was to facilitate these lessons-learnt to inform future responses. The themes identified then constitute the basis of a framework for future evaluations to ensure coherent and consistent learning in the organisation.

Source: ACF Emergency Response Learning Project (2011). Briefing note by (ACF-UK, 2011) Evaluations, Learning & Accountability Unit.

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. When evaluation findings and results are gathered from and synthesised across several evaluations, potential users of such information have more data points and evidence at their disposal, and may place greater confidence in the results presented to them. Greater consistency of findings across programmes also leads to more confidence in their credibility and to greater willingness to make changes based on them.

Box 17

Increasing use of evaluation findings in WFP

WFP has put significant resource and effort into improving the impact it gets from its evaluations by focusing on the post-evaluation process in order to increase access to and use of relevant and timely evidence from evaluations:

- It carried out syntheses of country-level reviews and evaluations to plug into country-level decision-making processes such as the development of new country strategies. The Ethiopia Country Synthesis, for example, drew on 16 evaluations and studies and presented key lessons from past experience in the same framework as that required for the forward-looking Country Strategy. All the lessons were referenced to the source reports, so further reading was facilitated. This has been popular with the country offices.
- WFP produced a synthesis of the 4 strategic evaluations of different dimensions of WFP's Strategic Plan 2008-2013. From this synthesis other tailor-made products were made, such as one for senior managers to use in a special strategy workshop. Products that synthesise - particularly for specific decision-making moments - are very popular.

- WFP produced a synthesis of the 4 strategic evaluations of different dimensions of WFP's Strategic Plan 2008-2013. From this synthesis other tailor-made products were made, such as one for senior managers to use in a special strategy workshop. Products that synthesise - particularly for specific decision-making moments - are very popular.
- Four themed 'top ten lessons' were prepared, based on the ALNAP model: on targeting, cash & vouchers, gender and on safety nets. Evaluation briefs have been prepared for all evaluation reports completed since 2011.
- Policy evaluations feed into the planning, monitoring and evaluation cycle of new and existing policies and Country Portfolio Evaluations are already timed to provide evidence for the preparation of WFP country strategies.
- Lessons from the set of school feeding impact evaluations were presented to a corporate consultation on school feeding, and to an international technical meeting on home-grown school-feeding. A lunchtime seminar for staff was also organised.
- Evaluations are accessible in the evaluation library on WFP's official website. A variety of products are available for drawing lessons from evaluations tailored to specific audiences.

Sources: Sally Burrows, WFP, personal communication, March 2013; WFP, 2009; 2011.

Concluding observations

In this paper, we have sought to highlight key issues around building the capacity to carry out and use – in the widest sense of the term – evaluation of humanitarian action (EHA). Drawing on published and grey literature, interviews and workshops with humanitarian evaluators, we have developed an evaluation capacities framework to facilitate agency self-analysis and to structure debate around the key issues identified in this study. We hope that agencies will be able to use the findings presented within the framework to improve their capacity to fund, commission, support, carry out, communicate, and meaningfully use humanitarian evaluations.

The framework is hierarchical, with the most important and fundamental issues of leadership, culture, structure and resources appearing in Capacity Area 1 'Leadership, culture, structure and resources'. Clarifying purpose, demand and strategy are also important but less significant and so appear in Capacity Area 2 'Purpose, demand and strategy'. Capacity Area 3 'Evaluation processes and systems' focuses on processes and systems that, while useful in their own right, are considered less likely to bring about fundamental change on their own, without changes made elsewhere.

It may be easier to implement the recommendations of Capacity Area 3 than those in other areas, and there may be gains that could be achieved here that will drive more fundamental change elsewhere. At the same time, the three Capacity Areas are mutually reinforcing. Some of the easy-wins from a 'lower' capacity area may help catalyse interest in evaluation, which can promote change in more challenging capacity areas. Small changes in one area may trickle up, and overflow into other capacity areas[32].

It seems fair to note that no one-size-fits-all approach could be derived from this exploration of humanitarian evaluation capacities. This is also because – as others have noted – increasing the impact of humanitarian evaluations is also about constructing pathways for the evaluation findings to make a difference within the organisation.

[32] Margie Buchanan-Smith, Independent, personal communication, June 2013.

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 is achieved (Stoddard, 2005).

We have populated the capacities framework for humanitarian evaluation presented in this paper with case studies and insights gained from the interviews and workshops, and through the ALNAP Evaluation Community of Practice. For some areas of the framework, it was more challenging to elicit practical examples and insights from practitioners. This difficulty perhaps reflects a failure to ask the right questions, or that evaluators and practitioners experience in these areas has not yet been captured and analysed more thoroughly.

What has been very apparent throughout this process is that evaluators struggle under a heavy workload. We have greatly appreciated the time that evaluators have found to engage with this project. Indeed, the energy and passion we encountered in the evaluators within the humanitarian sector is astonishing. Evaluation of humanitarian action remains a relatively new discipline, and is evolving all the time, but seems to be in good hands!

The framework we suggest in this study is deliberately light, to allow busy evaluators to access it easily and efficiently. It is not a step-by-step manual on how to implement change. Rather, we have designed it to motivate and inspire, and to serve as a basis for reflection. Currently, a step-by-step manual would not be possible, as evidence on the optimal sequencing of capacity-strengthening actions across the three components of the framework does not yet exist. However, many interviewees believe that leadership requires most attention initially. Our contributors also considered securing adequate resources to be very important. It seems likely, however, that there is no single approach that will work universally, as organisations in the sector are so very different in size, culture and ambition.

Hence, it is important for organisations to conduct their own internal analysis to conceive the most appropriate approach for improving evaluation utilisation. As a follow-up to this study, a self-assessment tool will be designed to help agencies reflect on their evaluation processes, take stock of their practice in evaluation utilisation and uptake, and identify areas on which to focus future efforts[33].

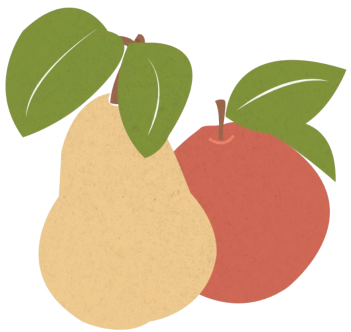
[33]The self-assessment tool will be made available on the ALNAP website at www.alnap.org/using-evaluation

A pilot version of the tool was tested with humanitarian practitioners, evaluators and those who have contributed to this study. Several organisations found it particularly useful when it was administered to different sections and branches of the organisation rather than just to the evaluation office, with the findings then being discussed jointly. Field staff and evaluators can differ markedly in their perception and assessment of their own organisation's evaluation capacity. The tool could help highlighting some of the perceptions and misperceptions that often gravitate around evaluation work, and could help encourage discussions on improving evaluation use.

There is a range of material available for further reading on building evaluation capacities, albeit not with a focus on the humanitarian sector. Much of this literature focuses on the supply-side, and advocates a range of structural, personnel, technical and financial change processes. We have designed this paper to be less formulaic in approach, and to focus as much on demand as on supply. Without stimulating demand for evaluation, there is less chance that structural change will happen.

Members of the humanitarian evaluation community will be able to continue the debates through the ALNAP Evaluation Community of Practice (www.partnerplatform.org/humanitarian-evaluation). More experience of attempts to improve evaluation capacities in the humanitarian sector will lead to richer insights and greater knowledge of what works. It is important that a mechanism to share this continues to exist.

It seems likely, however, that there is no single approach that will work universally, as organisations in the sector are so very different in size, culture and ambition.



List of contributors

This list includes all those who contributed to this study through either interviews or participation in the ALNAP Evaluation Community of Practice. The examples and insights they shared with us have greatly enriched this work and bring it closer to evaluation practitioners' realities.

Organisation	Contributors
ACF	Saul Guerrero, Ben Allen
American Red Cross	Dale Hill (now Independent)
CDA	Dayna Brown
DARA	Riccardo Polastro
DEC	Annie Devonport
DFID	Jonathan Patrick
Groupe URD	François Grünewald, Véronique de Geoffroy
Independent	John Cosgrave, Margie Buchanan-Smith, Jock Baker, Tony Beck , Abhijit Bhattacharjee, Carlisle Levine Monica Oliver
IRC	Anjuli Shivshanker
MSF	Sabine Kampmueller
Oxfam America	Silva Sedrakian
Oxfam GB	Vivien Walden
Oxfam Intermón	Luz Gómez Saavedra
Save the Children International	Dan Collison
Save the Children US	Hana Crowe
Sida	Joakim Molander (now Swedish Ministry of Foreign Affairs)
Tearfund	Catriona Dejean
UNHCR	Jeff Crisp
UNICEF	Rob McCouch (now UN OIOS) Erica Mattellone
UNRWA	Robert Stryk
WFP	Sally Burrows

Bibliography

ACF International (2011a) *Evaluation policy and guidelines: enhancing organisational practice through an integrated evaluations, learning and accountability framework.* London: ACF International. (www.alnap.org/resource/6199.aspx).

ACF International (2011b) *Learning Review 2011*. London: ACF International. (www.alnap.org/resource/7001.aspx).

ACF International (2013) 2012 *Learning Review.* London: ACF International. (www.alnap.org/resource/8219.aspx).

ACF-UK (2011) 'ACFIN emergency response learning project'. Briefing note. (internal document) London: Evaluations, Learning & Accountability Unit ACF-UK. (www.alnap.org/resource/8648.aspx).

Allen, M. (2010) *Steeping the organization's tea: examining the relationship between evaluation use, organizational context, and evaluator characteristics.* Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Mandel School of Applied Social Sciences, Case Western Reserve University. (www.alnap.org/resource/8532.aspx).

Apthorpe, R., Borton, J. and Wood, A. (eds) (2001) *Evaluating international humanitarian action: Reflections from Practitioners.* London: Zed Press (www.alnap.org/resource/5287.aspx).

ALNAP (2012) *The State of the Humanitarian System*. London: ALNAP/ODI. (www.alnap.org/sohsreport.aspx).

ALNAP and Groupe URD (2009) *Participation handbook for humanitarian field workers: involving crisis-affected people in a humanitarian response.* Plaisians: Groupe URD. (www.alnap.org/resource/8531.aspx).

Anderson, M., Brown, D. and Jean, I. (2012) *Time to listen: hearing people on the receiving end of international aid.* Cambridge, Massachusetts: CDA Collaborative Learning Projects. (www.alnap.org/resource/8530.aspx).

Baker. J., Christoplos, I., Dahlgren, S., Frueh, S., Kliest, T., Ofir, Z. and Sandison P. (2007) *Peer review: evaluation function at the World Food Programme.* Sida (www.alnap.org/resource/3604.aspx).

Bamberger, M. (2012) *Introduction to mixed methods in impact evaluation.* Impact Evaluation Notes. Washington DC: InterAction (www.alnap.org/resource/8079.aspx).

Beck, T. (2002) 'Meta-evaluation' in *ALNAP Annual Review: improving performance through improved learning.* London: ALNAP/ODI (www.alnap.org/resource/5194.aspx).

Beck, T. (2004a) 'Meta-Evaluation' in in *ALNAP Review of Humanitarian Action in 2003:* Field level learning. London: ALNAP/ODI (www.alnap.org/resource/5210.aspx).

Beck, T. (2004b) 'Learning and evaluation in humanitarian action', in *ALNAP Review of Humanitarian Action in 2003: Field level learning.* London: ALNAP/ODI (www.alnap.org/resource/5207.aspx).

Beck, T. (2006) 'Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies'. London: ALNAP/ODI (www.alnap.org/resource/5253.aspx).

Beck, T. and Buchanan-Smith, M. (2008) 'Joint evaluations coming of age? The quality and future scope of joint evaluations' in *ALNAP 7th Review of Humanitarian Action.* London: ALNAP/ODI (www.alnap.org/resource/5232.aspx).

Beck, T. and Buchanan-Smith, M. (2008) 'Meta-evaluation' in *ALNAP 7th Review of Humanitarian Action*. London: ALNAP/ODI (www.alnap.org/resource/8410.aspx).

Bonbright, D. (2012) 'Use of impact evaluation results'. Impact Evaluation Note 4. Washington, D.C.: InterAction. (http://www.alnap.org/resource/8534.aspx).

Bourgeois, I., Hart, R., Townsend, S. and Gagné, M. (2011) 'Using hybrid models to support the development of organizational evaluation capacity: A case narrative', *Evaluation and Program Planning* 34: 228–235. (www.alnap.org/resource/8389.aspx).

Boyle, R. and Lemaire, D. (eds) (1999) 'Building effective evaluation capacity: lessons from practice'. Comparative Policy Evaluation 8. New Brunswick: Transaction Publishers (www.alnap.org/resource/8501.aspx).

Buchanan-Smith, M. with Scriven, K. (2011) *Leadership in action: leading effectively in humanitarian operations*. London: ALNAP/ODI (www.alnap.org/resource/6118.aspx).

Catley, A., Burns, J., Abede, D. and Suji, O. (2008) *Participatory impact assessment: a guide for practitioners*. Medford, MD: Feinstein International Center, Tufts University. (www.alnap.org/resource/8094.aspx).

Carlsson, J., Koehlin, G. and Ekborn, A. (1994) *The political economy of evaluation: international aid agencies and the effectiveness of aid*. London: Macmillan Press. (www.alnap.org/resource/8411.aspx).

Chelimsky, E. (1997) 'The coming transformations in evaluation', in Chelimsky E. and Shadish, W. (eds), *Evaluation for the 21st century: A Handbook*. Thousand Oaks, CA: Sage (www.alnap.org/resource/8412.aspx).

CIDA (2012) *Annual report 2011–2012: CIDA learns – lessons from evaluations* (www.alnap.org/resource/8413.aspx).

Clarke, P. and Ramalingam, B. (2008) *Organisational change in the humanitarian sector*. London: ALNAP/ODI (www.alnap.org/resource/5231.aspx)

Cosgrave, J. (2008) *Responding to earthquakes: learning from earthquake relief and recovery operations*. ALNAP Lessons Paper. London: ALNAP/ODI (www.alnap.org/resource/5239.aspx).

Cosgrave, J. and Buchanan-Smith, M. (2013) *Evaluating humanitarian action: ALNAP pilot guide*. London: ALNAP/ODI (www.alnap.org/eha).

Cosgrave, J., Ramalingam, B. and Beck, T. (2009) *Real-time evaluations of humanitarian action – an ALNAP guide*. Pilot version. London: ALNAP/ODI (www.alnap.org/resource/5595.aspx).

Cosgrave, J., Polastro, R. and van Eekelen, W. (2012) *Evaluation of the Consortium of British Humanitarian Agencies (CBHA) pilot* (www.alnap.org/pool/files/1448.pdf).

Crisp, J. (2000) 'Thinking outside the box: evaluation and humanitarian action', in *Forced Migration Review* 8: 4-7 (www.alnap.org/resource/8400.aspx).

DAC/UNEG (2007) 'Framework for professional peer reviews'. DAC/UNEG Joint Task Force on Professional Peer Reviews of Evaluation Functions in Multilateral Organizations. New York: UNEG (www.alnap.org/resource/8401.aspx).

Darcy, J. (2005) 'Acts of faith? Thoughts on the effectiveness of humanitarian action'. Paper prepared for the Social Science Research Council seminar series, *The transformation of humanitarian action*, 12 April, New York. London: Overseas Development Institute. (www.alnap.org/resource/8414.aspx).

DEC (2013) 'DEC approach to learning'. Paper 7, February. (internal document) London: DEC. (www.alnap.org/resource/8649.aspx).

DFID (2011a) 'How to note: writing a business case'. DFID Practice Paper. (www.alnap.org/resource/8402.aspx).

DFID (2011b) 'Department for International Development organisation chart'. (www.alnap.org/resource/8403.aspx).

DFID (2013) *International development evaluation policy*. London: Department for International Development (www.alnap.org/resource/8404.aspx).

Drew, R. (2011) *Synthesis study of DFID's strategic evaluations 2005–2010*. Report produced for the Independent Commission for Aid Impact. (www.alnap.org/resource/8405.aspx).

ECB (2007) *Good enough guide: impact measurement and accountability in emergencies*. Warwickshire: Practical Action Publishing. (www.alnap.org/resource/8406.aspx).

ECB (2011) 'What we know about joint evaluations of humanitarian action: learning from NGO experiences'. Version 6, April. (www.alnap.org/resource/7911.aspx).

ECB - AIM Standing Team (2011) *Case study on Accountability and Impact Measurement Advisory Group* (www.alnap.org/resource/8407.aspx).

Engel, P., Carlsson, C. and van Zee, A. (2003) *Making evaluation results count: internalising evidence by learning.* Policy Management Brief 16 (www.alnap.org/resource/8408.aspx).

FAO (2012) *Peer Evaluation: The Evaluation Function of the Food and Agriculture Organization.* Final Report. Washington, DC: Global Facility Evaluation Office (www.alnap.org/resource/8505.aspx).

Fearon, J. (2005) 'Measuring humanitarian impact'. Discussion paper for the SSRC seminar series, *The Transformation of Humanitarian Action,* 12 April, New York. (www.alnap.org/resource/8415.aspx).

Feinstein, O. (2002) 'Use of evaluations and the evaluation of their use', *Evaluation* 8(4): 433–439 (www.alnap.org/resource/8416.aspx).

Foresti, M. (2007) 'A comparative study of evaluation policies and practices in development agencies'. Série Ex Post : notes méthodologiques 1, December. London: ODI. (www.alnap.org/resource/7773.aspx).

Forss, K., Krause, S.-E., Taut, S. and Tenden, E. (2006) 'Chasing a ghost? An essay on participatory evaluation and capacity development', *Evaluation* 1(6): 129 (www.alnap.org/resource/8417.aspx).

Frerks, G. and Hilhorst, D. (2002) 'Evaluation of humanitarian assistance in emergency situations'. Working Paper 56. New Issues in Refugee Research, UNHCR/EPAU (www.alnap.org/resource/8418.aspx).

Furubo, J.-E., Rist, R. and Sandahl, R. (eds) (2002) *International Atlas of Evaluation.* New Brunswick: Transaction Publishers. (www.alnap.org/resource/8536.aspx).

Government Office for Science (2011a) 'Science & engineering assurance review of the Department for International Development'. URN 11/1260 (www.alnap.org/resource/8419.aspx).

Government Office for Science (2011b) 'DIFD science & engineering assurance review: annex A – department background'. URN 11/1260 (www.alnap.org/resource/8419.aspx).

Grasso, P. (2003) 'What makes an evaluation useful? Reflections from experience in large organizations', *American Journal of Evaluation* 24(4): 507–514 (www.alnap.org/resource/8420.aspx).

Griekspoor A. and Sondorp, E. (2001) 'Enhancing the quality of humanitarian assistance: taking stock and future initiatives', in *Prehospital and Disaster Medicine* 16(4): 209–215 (www.alnap.org/resource/5322.aspx)

Groupe URD (2009) *The Quality Compass companion book.* Plaisians : Groupe URD (www.alnap.org/resource/8541.aspx).

Groupe URD (2010) *Étude en temps réel de la gestion de la crise en Haïti après le séisme du 12 janvier 2010.* Plaisians : Groupe URD (www.alnap.org/resource/8409.aspx).

Guba, E. and Lincoln, Y. (1989) *Fourth Generation Evaluation.* Newbury Pak, CA: Sage (www.alnap.org/resource/8421.aspx).

Guerrero, S. (2012) 'Learning to learn: harnessing external evaluations to build on strengths, enhance effectiveness'. ACF Blog, 14 August. London: Action Against Hunger. (www.alnap.org/resource/8537.aspx)

Guerrero, S. (2013) 'The needle in the haystack: the role of external evaluations in identifying and promoting good practices within NGOs'. Presentation prepared for the 28th ALNAP Meeting, Washington DC, 5-7 March. (www.alnap.org/resource/8538.aspx).

Guerrero, S., Woodhead, S. and Hounjet, M. (2013) *On the right track: a brief review of monitoring and evaluation in the humanitarian sector.* (www.alnap.org/resource/8211.aspx).

Hallam, A. (1998) 'Evaluating humanitarian assistance programmes in complex emergencies'. Relief and Rehabilitation Network, Good Practice Review 7. London: ODI. (www.alnap.org/resource/8210.aspx).

Hallam, A. (2011) 'Harnessing the power of evaluation in humanitarian action: an initiative to improve understanding and use of evaluation'. ALNAP Working Paper. London: ALNAP/ODI. (www.alnap.org/resource/6123.aspx).

Heider, C. (2011) 'Conceptual framework for developing evaluation capacities: building on good practice', in Rist, R., Boily, M.-H. and Martin, F. (eds), *Influencing Change:*

Evaluation and Capacity Building: 85–107. Washington DC: World Bank. (www.alnap.org/resource/8422.aspx)

House, E. (1983) 'Assumptions underlying evaluation models', in Madaus, G. F., Scriven, M. and Stufflebeam, D. L. (eds), *Evaluation Models.* Boston: Kluwer-Nijhoff (www.alnap.org/resource/8423.aspx).

Hovland, I. (2003) 'Knowledge management and organisational learning, an international development perspective: an annotated bibliography'. ODI Working Paper 224. London: ODI. (www.alnap.org/resource/8424.aspx).

IASC (2008) *Inter-agency RTE of the humanitarian response to cyclone Nargis in Myanmar.* Geneva and New York: IASC. (www.alnap.org/resource/8296.aspx).

IASC (2010) *Inter-agency real-time evaluation procedures and methodologies.* Geneva and New York: IASC. (www.alnap.org/resource/8425.aspx).

IASC (2011a) 'Inter-agency real-time evaluation of the humanitarian response to 2010 floods in Pakistan'. Geneva and New York: IASC. (www.alnap.org/resource/6087.aspx).

IASC (2011b) 'Inter-agency real-time evaluation of the humanitarian response to the earthquake in Haiti: 20 months after'. Geneva and New York: IASC. (www.alnap.org/resource/6330.aspx).

IASC (2012) *IASC transformative agenda – 2012.* Geneva and New York: IASC. (www.alnap.org/resource/8502.aspx).

ICAI (2011a) 'Independent Commission for Aid Impact – Work Plan'. (www.alnap.org/resource/8426.aspx).

ICAI (2011b) 'Independent Commission for Aid Impact: report to the House of Commons International Development Committee on ICAI's Shadow Period: November 2010 – May 2011'. (www.alnap.org/resource/8427.aspx).

ICAI (2012) 'Transparency, impact and Value for Money – The work of the ICAI and how it impacts INGOs'. Speech by Chief Commissioner at the Crowe Clark Whitehill INGO conference on 3 February 2012. (www.alnap.org/resource/8428.aspx).

IFRC (2011a) *IFRC framework of evaluation.* IFRC Secretariat: Planning and Evaluation Department (www.alnap.org/resource/7975.aspx).

IFRC (2011b) *Project/programme monitoring and evaluation (M&E): guide.* Geneva: IFRC. (www.alnap.org/resource/8542.aspx).

IFRC (2011c) *Revised plan 2011: disaster services – executive summary.* Geneva: IFRC. (www.alnap.org/resource/8468.aspx).

IFRC (2013) 'Learning from the Haiti operation: from a large-scale operation experience to a global and institutionalized knowledge'. Learning Concept Paper for Haiti. (internal document) Port-au-Prince: IFRC. (www.alnap.org/resource/8650.aspx).

Innovation Network (2013) *State of evaluation 2012: evaluation practice and capacity in the nonprofit sector.* (www.alnap.org/resource/8493.aspx).

Jones, N., Jones, H., Steer, L. and Datta, A. (2009) *Improving impact evaluation production and use.* Working Paper 300. London: ODI. (www.alnap.org/resource/8225.aspx).

Jones, H. and Mendizabal, E. (2010) *Strengthening learning from research and evaluation: going with the grain.* London: ODI (www.alnap.org/resource/8429.aspx).

Kellogg Foundation (2007) *Designing Initiative Evaluation.* Battle Creek, MI: WK Kellogg Foundation (www.alnap.org/resource/8469.aspx).

Lawrenz, F., Gullickson, A. and Toal, S. (2007) 'Dissemination: handmaiden to evaluation use', in *American Journal of Evaluation* 28(3): 275–89 (www.alnap.org/resource/8463.aspx).

Levine, C. and Chapoy, C. (2013) *Building on what works: Commitment to Evaluation (C2E) Indicator.* (www.alnap.org/resource/8319.aspx).

Lincoln, Y. (1991) 'The arts and sciences of program evaluation', in *Evaluation Practice* 12(1): 1 (www.alnap.org/resource/8464.aspx).

Lindahl, C. (1998) 'The management of disaster relief evaluation - Lessons from a Sida Evaluation of the Complex Emergency in Cambodia', in Sida Department for Evaluation and Internal Audit, Sida Studies in Evaluation. (www.alnap.org/resource/8551.aspx)

Lindahl, C. (2001) 'Evaluating SIDA's complex emergency assistance in Cambodia: conflicting perceptions', in Wood, Apthorpe and Borton (eds), *Evaluating International Humanitarian Action:* 59. London: Zed Books. (www.alnap.org/resource/8558.aspx)

Liverani, A. and Lundgren, H. (2007) 'Evaluation systems in development aid agencies: an analysis of DAC peer reviews 1996–2004', in *Evaluation* 13(2): 242 (www.alnap.org/resource/8467.aspx).

Mayne, J. (2008) 'Building an evaluative culture for effective evaluation and results management'. Institutional Learning and Change Initiative (ILAC) Brief 20, November (www.alnap.org/resource/8465.aspx).

Mayne J. and Rist, R. (2006) 'Studies are not enough: the necessary transformation of evaluation', in *Canadian Journal of Program Evaluation* 21(3): 93–120 (www.alnap.org/resource/8466.aspx).

Mackay, K. (2009) 'Building monitoring and evaluation systems to improve government performance' in *Country-led monitoring and evaluation systems - Better evidence, better policies, better development results.* Segone, M. (ed.) New York: UNICEF. 169-187 (www.alnap.org/resource/8567.aspx).

Mebrahtu, E., Pratt, B. and Lönqvist, L. (2007) *Rethinking Monitoring and Evaluation: Challenges and Prospects in the Changing Global Aid Environment.* Oxford: INTRAC (www.alnap.org/resource/8485.aspx).

McNulty, J. (2012) 'Symbolic uses of evaluation in the international aid sector: arguments for critical reflection', in *Evidence & Policy: a Journal of Research, Debate and Practice,* 8(4): 495-509 (www.alnap.org/resource/8497.aspx).

Minnett, A. M. (1999) 'Internal evaluation in a self-reflexive organization: One nonprofit agency's model', *Evaluation and Program Planning* (22): 353-362 (www.alnap.org/resource/8498.aspx).

Molund, S. and Schill, G. (2007) *Looking back, moving forward: Sida Evaluation Manual.* Sida (www.alnap.org/pool/files/evaluation_manual_sida.pdf).

Morras Imas, L. and Rist, R. (2009) *The road to results: designing and conducting effective development evaluations.* Washington DC: The World Bank (www.alnap.org/resource/8470.aspx).

Nelson, J. (2012) 'Actionable measurement at the Gates Foundation', Blog from 6 September 2012 (www.alnap.org/resource/8557.aspx).

NGO and Humanitarian Reform Project (2010) 'Good practices in humanitarian assistance: Zimbabwe'. Good Practice Paper Series (www.alnap.org/resource/8486.aspx).

OECD (2006) 'The Challenge of Capacity Development: Working Towards Good Practice'. DAC Guidelines and Reference Series (www.alnap.org/resource/8487.aspx).

OECD-DAC (2001) 'Evaluation feedback for effective learning and accountability'. Evaluation and Aid Effectiveness 5. Paris: OECD (www.alnap.org/resource/8488.aspx).

OECD-DAC (2002) *Glossary of Key Terms in Evaluation and Results-Based Management.* Paris: OECD/DAC (www.alnap.org/resource/8489.aspx).

Oliver, M. (2007) *CARE's humanitarian operations: review of CARE's use of evaluations and after action reviews in decision-making.* Georgia State University (www.alnap.org/resource/8058.aspx).

Oliver, M. (2009) 'Meta-evaluation as a means of examining evaluation influence', in *Journal of MultiDisciplinary Evaluation* 6(11): 32–37 (www.alnap.org/resource/8285.aspx).

O'Neil, G. (2012) '7 new ways to communicate findings'. Presentation at the European Evaluation Society Conference in Helsinki, Finland (www.alnap.org/resource/8471.aspx).

Oral History Project Team (2006) 'The oral history of evaluation, Part IV – the professional evolution of Carol H. Weiss', in *American Journal of Evaluation* 27(4): 478. (www.alnap.org/resource/8565.aspx)

Owen, J. and Lambert, F. (1995). 'Roles for evaluation in learning organisations', in Evaluation. I (2), 259-273 (www.alnap.org/resource/8494.aspx).

Oxfam GB (2012) *Project effectiveness reviews: summary of 2011/12 findings and lessons learned*. Oxford: Oxfam. (www.alnap.org/resource/8472.aspx).

Oxfam GB (n.d.) *How are effectiveness reviews carried out?* Oxford: Oxfam. (www.alnap.org/resource/8473.aspx).

Pankaj, V., Welsh, M. and Ostenso, L. (2011) 'Participatory analysis: expanding stakeholder involvement in evaluation'. Innovation Network. (www.alnap.org/resource/8553.aspx)

Patton, M.Q. (1990) 'The evaluator's responsibility for utilisation', in Alkin, M. (ed.) Debates on Evaluation. Newbury Park, CA: SAGE. (www.alnap.org/resource/8552.aspx)

Patton, M.Q. (1997) *Utilization-focused evaluation: the new century text*. 3rd ed. Thousand Oaks, CA: Sage (www.alnap.org/resource/8060.aspx).

Patton M.Q. (2001) 'Evaluation, knowledge management, best practices, and high. quality lessons learned', in *American Journal of Evaluation* 22(3) (www.alnap.org/resource/8477.aspx).

Patton, M.Q. (2010a) *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York: Guilford Press (www.alnap.org/resource/8499.aspx).

Patton, M.Q. (2010b) 'Incomplete successes', in *Canadian Journal of Program Evaluation* 25(3), Special issue. (www.alnap.org/resource/8651.aspx)

Pawson, R. and Tilley, N. (2004) *Realist evaluation*. Mount Torrens: Community Matters. (www.alnap.org/resource/8195.aspx).

Perrin, B. (2013) 'Evaluation of Payment by Results (PBR): Current Approaches, Future Needs'. Working Paper 39. London: DFID. (www.alnap.org/resource/8554.aspx).

Picciotto, R. (2007) 'The new environment for development evaluation', in *American Journal of Evaluation* 28(4): 521 (www.alnap.org/resource/8462.aspx).

Pratt, B. (2007) 'Rethinking monitoring and evaluation', in *Ontrac* 37 (September): 1 (www.alnap.org/resource/8461.aspx).

Preskill, H. and Torres, R. (1999) *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage (www.alnap.org/resource/8460.aspx).

Preskill, H. and Boyle, S. (2008) 'A multidisciplinary model of evaluation capacity building', in *American Journal of Evaluation* 29: 443–459 (www.alnap.org/resource/8459.aspx).

Proudlock, K. (2009) 'Strengthening evaluation capacity in the humanitarian sector: a summary of key issues' (author's unpublished notes).

Proudlock, K., Ramalingam, B. with Sandison, P. (2009) 'Improving humanitarian impact assessment: bridging theory and practice', in *ALNAP 8th Review of Humanitarian Action*. London: ALNAP/ODI (www.alnap.org/resource/5663.aspx).

Quesnel, J. S. (2006) 'The importance of evaluation associations and networks', in Segone, M. (ed), *New Trends in Development Evaluation*, Evaluation Working Papers, 5: 17-25. UNICEF, CEE/CIS, IPEN. (www.alnap.org/resource/8453.aspx).

Ramalingam, B. (2005) 'Implementing knowledge strategies: lessons from international development agencies'. ODI Working Paper 244. London: Overseas Development Institute. (www.alnap.org/resource/8457.aspx).

Ramalingam, B. (2006) *Tools for knowledge and learning: a guide for development and humanitarian organisations*. London: Overseas Development Institute. (www.alnap.org/resource/8456.aspx)

Ramalingam, B. (2011) 'Learning how to learn: eight lessons for impact evaluations that make a difference'. ODI Background Notes (April). London: Overseas Development Institute. (www.alnap.org/resource/8458.aspx).

Reed, E. and Morariu, J. (2010) 'State of evaluation: evaluation practice and capacity in the nonprofit sector'. Innovation Network (www.alnap.org/resource/8455.aspx).

Results Matter (2013) 'Humanitarian Genome Project: end of project evaluation, June 2013' (internal document) (www.alnap.org/resource/8652.aspx).

Rey, F. and Urgoiti, A. (2005) 'Manual de gestión del ciclo del proyecto en acción humanitaria'. Fundación la Caixa. (www.alnap.org/resource/8556.aspx).

Ridde, V. and Sahibullah, S. (2005) 'Evaluation capacity building and humanitarian organization', *Journal of MultiDisciplinary Evaluation* (JMDE) 3: 78–112 (www.alnap.org/resource/8478.aspx).

Riddell, R. (2009) 'The quality of DFID's evaluation reports and assurance systems: a report for ICAI based on the quality review' (internal document) London: IACDI (www.alnap.org/resource/8496.aspx).

Rist, R. (1990) *Program evaluation and the management of government: patterns and prospects across eight nations.* New Brunswick: Transaction Publishers (www.alnap.org/resource/8454.aspx).

Rist, R., Boily, M.-H. and Martin, F. (2011) 'Introduction. Evaluation capacity building: a conceptual framework', in Rist, R., Boily, M.-H. and Martin, F. (eds), *Influencing Change: Evaluation and Capacity Building*: 1–11. Washington, DC: World Bank. (www.alnap.org/resource/8555.aspx).

Rist, R., Boily, M.-H. and Martin, F. (2013) *Development evaluation in times of turbulence: dealing with crises that endanger our future.* Washington, DC: World Bank (www.alnap.org/resource/8476.aspx).

Rogers, P. (2009) 'Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation in designing impact evaluations: different perspectives'. 3ie Working Paper 4. London: 3ie (www.alnap.org/resource/8474.aspx).

Rossi, P. and Freeman, H. (1993) *Evaluation: a systematic approach.* 5th ed. Newbury Park, CA: Sage (www.alnap.org/resource/8500.aspx).

Rugh, J. and Segone, M. (eds) (2013) 'Voluntary Organizations for Professional Evaluation (VOPEs) learning from Africa, Americas, Asia, Australasia, Europe and Middle East'. EvalPartners, UNICEF, IOCE. (www.alnap.org/resource/8475.aspx).

Sandison, P. (2006) 'The utilisation of evaluations', *ALNAP Review of Humanitarian Action: Evaluation Utilisation.* London: ALNAP/ODI (www.alnap.org/resource/5288.aspx).

SCHR (2010) *SCHR peer review on accountability to disaster-affected populations.* Geneva: SCHR (www.alnap.org/pool/files/schr-peer-review-lessons-paper-january-2010.pdf).

Scriven, M. (1991) *Evaluation thesaurus.* 4th ed. Newbury Park, CA: Sage (www.alnap.org/resource/8480.aspx).

Segone M. and Ocampo, A. (eds) (2007) *Creating and developing evaluation organizations - lessons learned from Africa, Americas, Asia, Australasia and Europe.* Lima: IOCE, desco and UNICEF (www.alnap.org/resource/8503.aspx).

Spilsbury, M.J., Perch, C. , Norgbey, S., Rauniyar, G. and Battaglino, C. (2007) 'Lessons learned from evaluation: a platform for sharing knowledge'. Special Study Paper Number 2. Nairobi: Evaluation and Oversight Unit, United Nations Environment Programme (www.alnap.org/resource/8452.aspx).

Start, D. and Hovland, I. (2004) 'Tools for policy impact: a handbook for researchers'. Toolkit. London: ODI (www.alnap.org/resource/8479.aspx).

Stern, E. et al. (2012) 'Broadening the range of designs and methods for impact evaluations'. DFID Working Paper 38 (www.alnap.org/resource/8196.aspx).

Stoddard, A. (2005) *Review of UNICEF's evaluation of humanitarian action.* New York: United Nations Children's Fund (www.alnap.org/resource/8481.aspx).

Stufflebeam, D. (2000) 'Lessons in contracting for evaluation', in *American Journal of Evaluation* 21: 293–314 (www.alnap.org/resource/8482.aspx).

Taut, S. (2007) 'Defining evaluation capacity building: utility considerations', in *American Journal of Evaluation* 28(1) (www.alnap.org/resource/8483.aspx).

Tavistock Institute (2003) *The evaluation of socio-economic development. The Guide.* (www.alnap.org/resource/8484.aspx).

Tearfund (2013) 'Impact, monitoring and evaluation planning and tracking tool' (internal document) Teddington: Tearfund (www.alnap.org/resource/8653.aspx).

Torres, R.T. (1991) 'Improving the quality of internal evaluation: the evaluator as consultant-mediator', in *Evaluation and Program Planning* 14: 189–198 (www.alnap.org/resource/8492.aspx).

UNEG (2005) *Norms for evaluation in the UN system.* New York: UNEG (www.alnap.org/resource/8215.aspx).

UNEG (2008) *UNEG Code of Conduct for Evaluation in the UN system.* New York: UNEG (www.alnap.org/resource/8445.aspx).

UNEG (2010) *UNEG good practice guidelines for follow up to evaluations.* New York: UNEG (www.alnap.org/resource/8446.aspx).

UNEG (2011) *UNEG framework for professional peer reviews of the evaluation function of UN organizations.* Reference Document. New York: UNEG (www.alnap.org/resource/8447.aspx).

UNEG (2012a) 'National evaluation capacities'. Proceedings from the 2nd International Conference, 12–14 September 2011. UNEG Members' Publications (www.alnap.org/resource/8448.aspx).

UNEG (2012b) *Evaluation capacity in the UN System.* New York: UNEG (www.alnap.org/resource/8216.aspx).

UNEG (2012c) 'National evaluation capacity development: practical tips on how to strengthen National Evaluation Systems'. New York: UNEG (www.alnap.org/resource/7477.aspxs).

UNEG (2012d) 'Update note on peer reviews of evaluation in UN organizations'. UNEG-DAC Peer Reviews. New York: UNEG (www.alnap.org/resource/8449.aspx).

UNEG (2013) 'Lessons learned study of peer reviews of UNEG evaluation functions'. UNEG-DAC Peer Reviews. New York: UNEG (www.alnap.org/resource/8450.aspx).

UNEG/MFA (UNEG and Evaluation Department, Ministry of Foreign Affairs of Denmark) (2010) 'Peer review of the evaluation function at UNIDO'. New York: UNEG (www.alnap.org/resource/8451.aspx).

UNICEF (2006) 'New trends in development evaluation'. Segone, M. (ed.). Evaluation Working Papers 5. Geneva: UNICEF CEE/CIS, Regional Office for Central and Eastern Europe and the CIS Countries (www.alnap.org/resource/8453.aspx).

UNICEF (2010a) *Bridging the gap: the role of monitoring and evaluation in evidence-based policy-making.* New York: UNICEF (www.alnap.org/resource/8443.aspx).

UNICEF (2010b) 'Evaluation of the UNICEF education programme in Timor Leste: 2003-2009'. New York: UNICEF (www.alnap.org/resource/6383.aspx).

UNICEF (2010c) 'Global Evaluation Reports Oversight System (GEROS)'. New York: UNICEF (www.alnap.org/resource/7948.aspx).

UNICEF (2012a) 'Evaluation of UNICEF's Cluster Lead Agency Role in Humanitarian Action (CLARE): Cluster Analysis Summary Matrix' (internal document) New York: UNICEF (www.alnap.org/resource/8268.aspx).

UNICEF (2012b) 'Evaluation of UNICEF's Cluster Lead Agency Role in Humanitarian Action (CLARE): country case study selection – technical note' (internal document) New York: UNICEF (www.alnap.org/resource/8268.aspx).

UNICEF (2012c) 'Evaluation of UNICEF's Cluster Lead Agency Role in Humanitarian Action (CLARE): terms of reference'. Final Draft, 22 October (www.alnap.org/resource/8268.aspx).

UNICEF (2013) 'Thematic synthesis report on evaluation of humanitarian action'. Discussion paper submitted to the Executive Board for UNICEF's Annual Session 2013. (www.alnap.org/resource/8442.aspx).

UN General Assembly (2012) Resolution 67/226 on the Quadrennial Comprehensive Policy Review, adopted 21 December 2012. (www.un.org/en/development/desa/oesc/qcpr.shtml).

USAID (2009) 'Trends in international development evaluation theory, policy and practices'. (www.alnap.org/resource/8441.aspx).

Vogel, I. (2012) 'Review of the use of "theory of change" in international development. London: DFID (www.alnap.org/resource/7434.aspx).

Wagenaar, S. (2009) 'Peer review on partnership in humanitarian aid: an approach for mutual learning and cooperation between organisations (North and South)'. Wageningen: PSO (www.alnap.org/resource/8440.aspx).

Webster, M. and Walker, P. (2009) *One for all and all for one: Intra-organizational Dynamics in humanitarian action.* Medford, MA: Feinstein International Center (www.alnap.org/resource/8439.aspx).

Weiss, C. (1983a) 'Ideology, interests, and information: the basis of policy positions', in Callahan, D. and Jennings, B. (1983) *Ethics, the Social Sciences, and Policy Analysis.* New York: Plenum. (www.alnap.org/resource/8539.aspx)

Weiss, C. (1983b) 'The stakeholder approach to evaluation: origins and promise', in Anthony Bryk (ed.), *New Directions for Programme Evaluation: Stakeholder-Based Evaluation*: 3-14. New Direction for Programme Evaluation Series 17. San Francisco: Jossey Bass. (www.alnap.org/resource/8438.aspx)

Weiss, C. (1990) 'Evaluation for decisions: is anybody there? Does anybody care?', in M. Alkin (ed.), *Debates on Evaluation:* 171–184. Newbury Park, CA: Sage (www.alnap.org/resource/8437.aspx).

Weiss, C. (1998) *Evaluation – Methods for studying programs and policies* (2nd edition). NJ: Prentice Hall. (http://www.alnap.org/resource/8540.aspx)

WFP (2008a) *Evaluation of WFP's capacity development policy and operations.* Report. Rome: Office of Evaluation, World Food Programme (www.alnap.org/resource/8490.aspx).

WFP (2008b) *WFP evaluation policy* (www.alnap.org/resource/8491.aspx).

WFP (2009) *Closing the learning loop – harvesting lessons from evaluations: report of phase I.* Rome: Office of Evaluation, World Food Programme (www.alnap.org/resource/8436.aspx).

WFP-OE (2012) *Annual evaluation report: measuring results, sharing lessons.* (www.alnap.org/resource/8435.aspx).

Wiles, P. (2005) 'Meta-evaluation' in *ALNAP Review of Humanitarian Action in 2004: capacity building.* London: ALNAP/ODI (www.alnap.org/resource/5217.aspx).

Wood, A., Apthorpe, R. and Borton, J. (eds) (2001) *Evaluating international humanitarian action: reflections from practitioners.* London: Zed Books (www.alnap.org/resource/5287.aspx).

World Bank (2009) 'Making smart policy: using impact evaluation for policy making'. Doing impact evaluation 14 (www.alnap.org/resource/8434.aspx).

WB–IEG (2009) *Institutionalising impact evaluation within the framework of a monitoring and evaluation system.* Washington DC: World Bank Independent Evaluation Group (www.alnap.org/resource/8433.aspx).

York, N. (2012) 'Results, value for money and evaluation – a view from DFID'. Presentation delivered at the Development evaluation days, organized by AECID. Madrid 7 December 2012 (www.alnap.org/resource/8432.aspx).

Young, J. and Mendizabal, E. (2009) 'Helping researchers become policy entrepreneurs'. ODI Briefing Paper 53. London: ODI (www.alnap.org/resource/8431.aspx).

Young, J. and Court, J. (2004) 'Bridging research and policy in international development: an analytical and practical framework'. RAPID Briefing Paper. London: ODI. (www.alnap.org/resource/8430.aspx).

Related ALNAP publications

Also available at www.alnap.org/publications

Evaluating Humanitarian Action: ALNAP pilot guide

Harnessing the power of evaluation in humanitarian action: An initiative to improve understanding and use of evaluation

Improving humanitarian impact assessment: Bridging theory and practice

Joint evaluations coming of age? The quality and future scope of joint evaluations

Evaluating humanitarian action using the OECD-DAC criteria

The utilisation of evaluations

ALNAP
Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ
United Kingdom
Tel: + 44 (0)20 3327 6578
Fax: + 44 (0)20 7922 0399
Email: alnap@alnap.org

ALNAP would like to acknowledge the financial support of USAID in carrying out this initiative.