DIPLOMARBEIT

Titel der Diplomarbeit

„Evaluating Humanitarian Response to Disasters in the WASH Sector: An Analysis of the Benefits and Shortcomings in Using the OECD-DAC Criteria“

Verfasser

Mag.rer.nat. Philipp Polanski

angestrebter akademischer Grad

Magister (Mag.)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 057 390
Studienrichtung lt. Studienblatt:	Individuelles Diplomstudium Internationale Entwicklung
Betreuerin:	Dr. Elisabeth Ringhofer

Table of Contents

List of Figures

List of Tables

Abbreviations

ACF	Action Contre la Faim
ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
CAFOD	Catholic Agency for Overseas Development
DRR	Disaster Risk Reduction
EHA	Evaluation of Humanitarian Action
ECHO	Humanitarian Aid and Civil Protection Department of the European Commission
GWC	Global WASH Cluster
HAP	Humanitarian Accountability Partnership
HP	Hygiene Promotion
IFRC	International Federation of Red Cross and Red Crescent Societies
(I)NGO	(International) Non-governmental Organisation
LRRD	Linking Relief, Rehabilitation and Development
M&E	Monitoring and Evaluation
OECD-DAC	Organisation for Economic Co-operation and Development - Development Assistance Committee
RTE	Real-time Evaluation
UNICEF	United Nations Children's Fund
WASH	Water, Sanitation and Hygiene (Promotion)

1 Introduction

1.1 Background and Problem Statement

Humanitarian response to disasters is increasingly seen and assessed with its nexus to development. In the past years, a number of high-profile disasters (such as in Haiti, Pakistan or recently the Philippines) have highlighted the vulnerability of millions of people, and that it is far from transitory. People's resilience and coping strategies are embedded in their socio-economic and political positioning in society. As such, disasters which trigger humanitarian response are symptomatic of poverty and political crises. Linking relief, recovery and development remains challenging and for the most part unsolved. This is particularly evident in the field of *Water, Sanitation and Hygiene (WASH)*. A well-known shortcoming of many WASH emergency response operations is, that the strong focus on urgent needs tends to ignore taking into account and planning for the long-term needs and perspectives. Moreover, for both, the humanitarian and development sector, it is not unusual (for a great variety of reasons though), that after the end of a programme, water and sanitation provision are gradually declining.

However, it is not all about money or simply a matter of extending programmes. A shifting focus to the impact of humanitarian interventions has brought a renewed debate on the sector's performance. There is probably no other business characterised by such a strong commitment to "doing better". Since the early 1990s, dozens of quality and accountability initiatives have been developed by groups of humanitarian agencies to overcome poor performance. And within the humanitarian community there is common agreement that actors should be accountable to multiple stakeholders and that lessons learned should inform future action. A recurring theme in this regard is better engagement with disaster-affected people and communities; e.g. the *Active Learning Network for Accountability and Performance in Humanitarian Action's (ALNAP)* annual meeting in 2014 is dedicated to this old but even more crucial matter. The meeting's concept note (accessible through the ALNAP website under *http://www.alnap.org/meeting2014*) reasons the following:

"We've grasped the reality that better engagement might mean more relevant and efficient emergency responses. It also could lead to people being treated with dignity and preserving a sense of agency. Yet, despite this growing realisation, the insights gained from many ALNAP Members with experience in involving crisis-affected people in programming are scattered and often go unrecorded. This means the analysis and potential for sharing learning are lost.".

The latter points towards the specific problem this research work tackles out of the outlined complex of problems. But how do humanitarian agencies build knowledge on *What works, what not and why?* The primary source of evidence in humanitarian action stems from evaluation, unfortunately. It is so, simply because (in the absence of any other feasible and ethical research method) evaluations are the only systematic way to capture relevant insights of emergency response in a unique context with scientific methods. And unfortunately this is so, because many humanitarians feel that evaluations are not delivering their main purpose, i.e. informing future action and improving performance. The reasons for "evaluation ineffectiveness" within the humanitarian sphere are important, also well researched and therefore do not form a part of the thesis' research. The research interest of this work goes further back in the evaluation process and looks at the evaluation criteria itself, i.e. which dimensions of interventions should be evaluated, or in other words, what guides evaluators when addressing the questions an evaluation seeks to answer.

As disasters are complex, unique and dynamic phenomena, and evaluations are not a tool to capture everything, the lenses through which evaluations look at things are critical. The underlying criteria used in evaluations are ideally reflecting areas of weakness to provide pointers for better performance. The evaluation criteria developed by the *Development Assistance Committee* of the *Organisation for Economic Co-operation and Development (OECD-DAC)* have become prevalent in the *evaluation of humanitarian action (EHA)*. The OECD-DAC criteria are intended to be a comprehensive and complementary set of measures. Adapted from the evaluation of development initiatives to suit the context of disasters, these are *relevance/appropriateness, connectedness, coherence, coverage, efficiency, effectiveness and impact*. In addition, these criteria are supposed to be applied in combination with cross-cutting themes and humanitarian as well as sectoral standards.

Against the backdrop of the challenges in performance outlined above, the thesis explores the OECD-DAC criteria for evaluating WASH interventions in humanitarian response to disasters. In the focus are their recent practice, what they are used to reveal and if that is addressing persistent issues in the sector to create an evidence base for learning.

1.2 Research Questions

The diploma thesis examines three research questions to discuss the OECD-DAC criteria's usefulness for evaluating WASH emergency interventions in the light of identified areas of weakness:

- How are the OECD-DAC criteria used to evaluate humanitarian emergency response in the WASH sector and what do they reveal?
- How are cross-cutting themes and standards used in conjunction with the OECD-DAC criteria?
- What are the benefits and shortcomings of using the OECD-DAC criteria for evaluating WASH emergency response activities?

1.3 State of the Art

An ever growing range of literature on EHA themes is made available through various quality, accountability and learning initiatives in the humanitarian sector. In particular ALNAP has become one of the main hubs for evaluative resources and research. As an established set of evaluation criteria, the OECD-DAC criteria for EHA or rather selected aspects of them are frequently researched in the light of recent developments in humanitarian work. Such research work includes e.g. *Proudlock, K. and Ramalingam, B., 2009. Improving humanitarian impact assessment: bridging theory and practice. In: ALNAP 2009. ALNAP 8th Review of Humanitarian Action: Performance, Impact and Innovation. London: Overseas Development Institute.* Also, related research as undertaken in the thesis which critically reflects emerging issues when using the OECD-DAC for EHA has been done. An older example provides e.g. *Hallam, A., 1998. Evaluating Humanitarian Assistance Programmes in Complex Emergencies. London: Overseas Development Institute*. But like this study, available similar examinations do not focus on WASH. Furthermore, several EHA guides for evaluators and evaluation managers discussing the use of the OECD-DAC criteria amongst others are available. A recent exemplar is *Buchanan-Smith, M. and Cosgrave, J., 2013. Evaluation of Humanitarian Action (EHA). Pilot Guide.*

Occasionally, the use of OECD-DAC criteria in evaluations do also spur more sophisticated debates on their enhancements amongst researchers in respective academic journals; e.g. *Chianca, T., 2008. The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. Journal of MultiDisciplinary Evaluation, 5(9), pp.41-51.* and *Eggers, H.W., 2009. Comments and Proposals Concerning Chianca's "The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement". Journal of MultiDisciplinary Evaluation, 6(11), pp.116-124.* On the whole, the available literature indicates that existing research on EHA is rather generic. So far, single sectors of humanitarian work such as WASH have attracted limited attention in relation to EHA, while at the same time learning-oriented evaluations are done with a thematic scope and the calls for sector-wide evaluations instead of single agency evaluations are getting louder. In this sense, the thesis can be perceived as one response to the growing awareness that evaluations hold far more potential for learning within a sector than realised.

1.4 Definitions and Own Theoretical Position

Next to learning and informing future action, accountability is another prominent reason for doing evaluations. The thesis researches EHA as *"the systematic and objective examination of humanitarian action, intended to draw lessons to improve policy and practice and enhance accountability"* (Buchanan-Smith and Cosgrave, 2013, p.14). Based on the comprehensive definition by the *Good Humanitarian Donorship (GHD) Initiative*, humanitarian action is considered as listed below (GHD Initiative, 2003):

- *The objectives of humanitarian action are to save lives, alleviate suffering and maintain human dignity during and in the aftermath of man-made crises and natural disasters, as well as to prevent and strengthen preparedness for the occurrence of such situations.*
- *Humanitarian action should be guided by the humanitarian principles of humanity, meaning the centrality of saving human lives and alleviating suffering wherever it is found; impartiality, meaning the implementation of actions solely on the basis of need, without discrimination between or within affected populations; neutrality, meaning that humanitarian action must not favour any side in an armed conflict or other dispute where such action is carried out; and independence, meaning the autonomy of humanitarian objectives from the political, economic, military or other objectives that*

any actor may hold with regard to areas where humanitarian action is being implemented.

- *Humanitarian action includes the protection of civilians and those no longer taking part in hostilities, and the provision of food, water and sanitation, shelter, health services and other items of assistance, undertaken for the benefit of affected people and to facilitate the return to normal lives and livelihoods.*

In the context of the thesis, humanitarian action is also referred to as disaster response or emergency response. It is important to note that the distinction between humanitarian and development aid in the aftermath of a disaster is artificial and problematic because "it is often unclear when the emergency ends and recovery begins" (Buchanan-Smith and Cosgrave, 2013, p.13). Therefore, the term *relief to development contiguum* has been coined. Moreover, as Stokke (2007, p.5) states, "a key issue in humanitarian response concerns precisely the need to link humanitarian relief and rehabilitation to development and vulnerability reduction". One major aspect in linking relief and development encompasses what is framed with the term accountability. The latest *Humanitarian Accountability Report* argues that being accountable to affected people is "not just the right thing to do", but the "best way to ensure programmes are relevant, effective, efficient and sustainable" (Darcy, Alexander and Kiani, 2013, p.1).

Accountability is probably one of the most misconceived terms and concepts in the humanitarian community. At this point, a definition might not clarify what accountability may mean and entail, but shall be provided to indicate the background of how it is employed in the research. The *Humanitarian Accountability Partnership (HAP)* defines accountability as (HAP International, 2013, p.17):

"The means through which power is used responsibly. It is a process of taking account of, and being held accountable by, different stakeholders, and primarily those who are affected by the exercise of power.".

Within the humanitarian aid and development nexus, two aspects of accountability are considered essential. First, humanitarian action should strengthen, not undermine the accountability of government authorities to crisis-affected communities. And secondly, where this is not possible (e.g. in conflict contexts), humanitarian actors should also be accountable for strategic and policy decisions towards the people aimed to serve, rather than just short-term operational outcomes. After all, as Darcy, Alexander and Kiani (2013, p.17) assert, the "ultimate test of any scheme of accountability is improved outcomes for crisis-affected people, and it is on this basis that any proposal should ultimately be judged".

Finally, because theory of change models are hardly applicable in emergency response, deriving lessons from field experience gained through trial and error is inevitable. Accordingly, the best possible utilisation of evaluations would be to capture this process of testing. More evidence on *What works, what not and why?* can foster a rule of thumb for WASH in emergencies, leading to more relevant programmes with better quality and overall performance.

1.5 Scope and Aims

The thesis researches evaluations of WASH emergency response operations. It focuses on one specific aspect of evaluations, the evaluation criteria, and aims at:

- Identifying and highlighting key areas of weakness in the WASH sector
- Exploring how the OECD-DAC criteria are applied in recent evaluations
- Determining their usefulness as well as their blind spots
- Analysing the role of cross-cutting themes and standards in WASH evaluations
- Providing pointers for a more targeted use of the OECD-DAC criteria and how to enhance evaluations in the WASH sector

The intended audience of the research is the wider evaluation community, particularly evaluators and evaluation managers exposed to WASH. In elaborating a view on the foundations of EHA, the thesis' findings are best perceived as a starting point for work in progress. They shall spur on-going attempts of promoting accountability, learning and knowledge sharing in the WASH sector.

1.6 Methodology

At the heart of the research is a meta-analysis of evaluation reports. It is complemented by literature reviews, documentary research and a review of existing EHA approaches. The different parts of the methodology are detailed below:

- Meta-analysis of evaluation reports from WASH emergency response operations:

  Final reports from ten evaluations (listed in a separate section in the references, *Chapter 6*) have been examined based on the assessment matrix in the *ALNAP Quality Proforma* (ALNAP, 2005), attached in *Annex 1*. The meta-analysis particularly draws on one section of the matrix, i.e. *Section 4: Assessing the Intervention*, as not all of them are equally relevant to the research. This section offers a framework for assessing essential subareas of any emergency intervention: *Institutional Considerations; Needs Assessment, Objectives, Planning and Implementation; Application of EHA Criteria; Consideration given to Cross-cutting Issues.* The analysis informed by these assessment areas provides the qualitative data for the main part of the thesis.

  The full assessment matrix of the *ALNAP Quality Proforma* has been used together with other criteria as means for selecting evaluation reports from public accessible databases to reduce agency/context specific biases and ensure a data source with best possible analytical quality. The selection criteria were:

  - Focus on WASH emergency response
  - Diversity of contexts/emergencies
  - Evaluating recent interventions of major humanitarian organisations providing both, emergency and developmental WASH programmes

- Satisfactory compliance with the *ALNAP Quality Proforma* guidance on good evaluation practice

It is worth noting that only few evaluation reports on WASH in disasters are available in general. For the time being, the most prominent source for those is ALNAP's evaluative reports database, followed by sparse resources of some humanitarian actors and donors. Out of a few tens of potential evaluation reports, ten (actually eleven, because the evaluation of Barham, Nabunny and Philpott (2011a; 2011b) has produced two linked documents which could have been one if not required otherwise) were finally chosen, reflecting eight different emergency contexts/countries (*Pakistan, Liberia, The Democratic Republic of the Congo (DRC), South Sudan, Haiti, Maldives, Uganda and Zimbabwe*), and three WASH emergency actors (*Action Contre la Faim (ACF), Catholic Agency for Overseas Development (CAFOD) and the International Federation of Red Cross and Red Crescent Societies (IFRC)*) and one donor (*The Humanitarian Aid and Civil Protection Department of the European Commission (ECHO)*) as commissioning organisations. The selected evaluations also include two *Real-time Evaluations (RTE)* (Grünewald, et al., 2011a; Fortune and Rasal, 2010), a relatively new and in terms of learning promising EHA approach.

ACF, CAFOD and the IFRC are well established actors in the field of WASH and beyond that, committed to various quality and accountability initiatives. Therefore, they provide a qualified "sample" of implementing agencies for the thesis' research (ACF evaluation reports: Laurens, 2005; Luqman, Zulqarnain and Bahadar, 2012; CAFOD evaluation reports: Jeene, 2010; Van der Wijk, Mwezi and Kazan, 2010; IFRC evaluation reports: Fortune and Rasal, 2010; Fox, 2008; Grayel and Mattson, 2012). In addition, ECHO represents a major donor perspective in humanitarian response pursuing WASH together with *Linking Relief, Rehabilitation and Development (LRRD)*. The ECHO commissioned evaluation reports (Barham, Nabunny and Philpott, 2011a; Barham, Nabunny and Philpott, 2011b; DeVillez, Bousquet and Nyagwambo, 2011; Grünewald, et al., 2011a) of funded WASH interventions and their implementing organisations are considered to complement the view of the other evaluation reports commissioned by implementing actors.

- Review of academic literature on EHA and the use of the OECD-DAC criteria
- Analysis of documents referencing EHA guidelines and approaches as well as handbooks and standards for WASH programmes in disaster response

1.7 Overview

The next chapter starts with a short guide to WASH emergency response and key areas of weakness are identified. Followed by outlining the role of EHA and its application in the field of WASH, this chapter further introduces to the rationale of the OECD-DAC criteria and relevant quality and accountability standards. *Chapter 3* and *Chapter 4* then provide the main part of the research – an analysis of the evaluation reports linked with literature/document review. In *Chapter 3*, the OECD-DAC criteria and cross-cutting evaluation themes are examined. The intended use of the criteria and themes is contrasted with their applied use and extracted key findings are discussed. Similarly in *Chapter 4*, the gaps between theory and praxis of quality and accountability standards are pinpointed and major considerations guiding the research are elaborated. *Chapter 5* concludes with a summary of the main findings, answering the research questions as well as a critical review followed by potential prospects.

2 Setting the Scene – WASH, Evaluations and Standards in Humanitarian Response to Disasters

The first chapter of the thesis' main part sets out the background of WASH, EHA, evaluation criteria and standards in humanitarian work as the basis for the subsequent analysis in *Chapter 3 and 4*. It starts with a brief introduction to the state of the humanitarian WASH sector and what WASH in emergencies is all about. Furthermore, major gaps in WASH will be identified. The second section then draws attention to the conception of EHA as well as important issues to consider for evaluating emergency response and gives a first insight into the evaluation practice of WASH interventions in the selected evaluation reports. The third section explains the role of the OECD-DAC criteria for EHA and provides a summary of the concerns coming along with their application. In the chapter's last part, the quality and accountability elements of the Sphere and HAP standards are presented to exhibit their intended use.

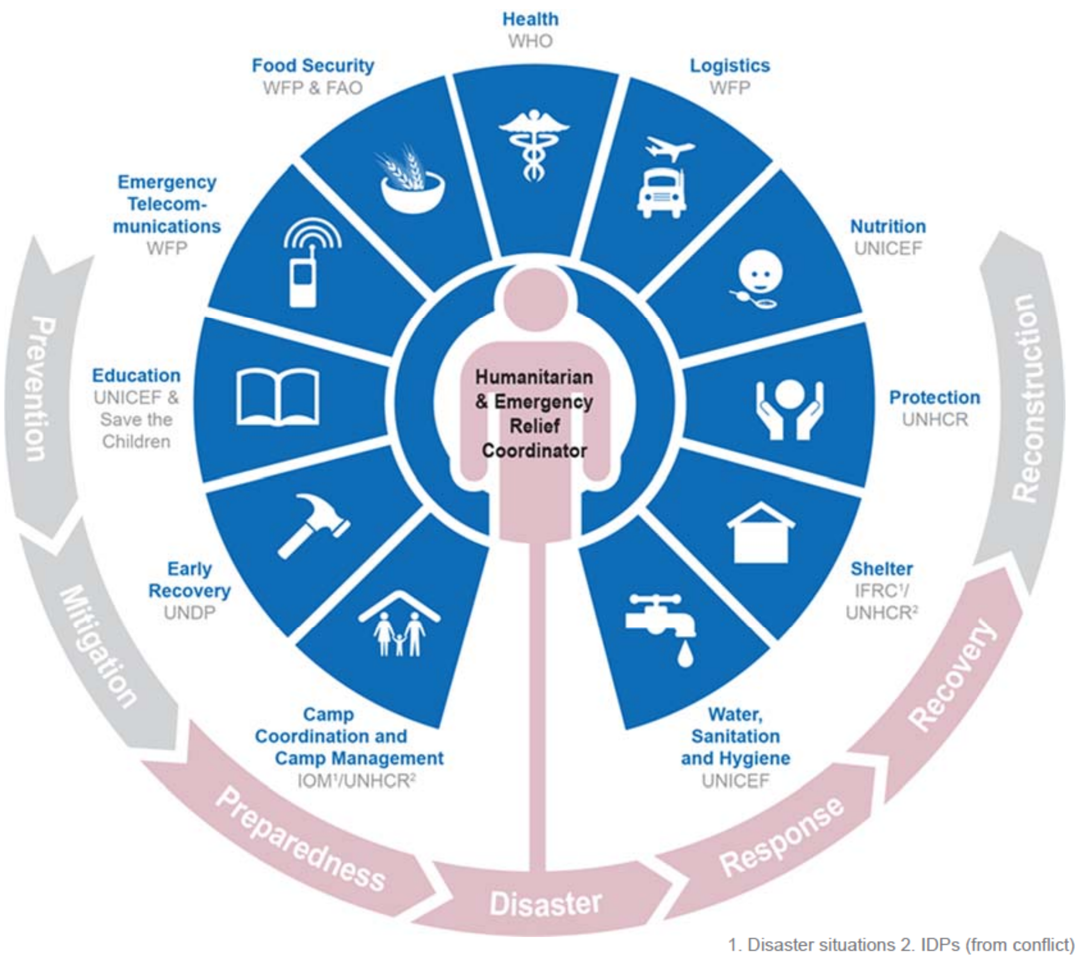
2.1 WASH Emergency Response and Current Challenges in a Nutshell

Despite the introductory notion of some felt and observed "evaluation ineffectiveness", it has to be acknowledged too, that past evaluations have indeed sparked substantial change of humanitarian action and the system it is embedded in. In particular the *Joint Evaluation of Emergency Assistance to Rwanda (JEEAR)* after the 1994 genocide and ensuing relief operations, and a decade later, the *Tsunami Evaluation Coalition (TEC)* in the aftermath of the Indian Ocean earthquake and tsunamis have had sweeping impact. Evidence from these evaluations initiated an external regulatory environment for humanitarian action (e.g. standards) and informed reforming humanitarian coordination.

The latter represented a long standing issue and one of the most extensive innovations humanitarian reform has induced, was the *Cluster Approach*. Clusters are groups of humanitarian organisations in each of the main sectors of humanitarian response such as WASH. An underlying idea is to foster partnerships between the various organisations, national/local authorities and civil society. Clusters have clear responsibilities for coordination and constitute a platform for implementing organisations and other major stakeholders throughout all phases of disaster management, i.e. *prevention, mitigation, preparedness, response, recovery and reconstruction*. *Figure 1* on the next page illustrates the cluster system.

Depending on the scale of an emergency, humanitarian needs and national response and coordination capacity, cluster activation can be requested as part of an international emergency response. The first time this happened was in 2005, following an earthquake in Pakistan. Since then, according to data from the clusters (accessible through the *humanitarianresponse.info* web portal under *https://www.humanitarianresponse.info*), the cluster approach has been used in over 30 countries. It is intended to add value to humanitarian coordination through (ibid.):

- *Increased transparency and accountability*
- *Enhanced predictability*
- *Engagement with national and local authorities*
- *Inclusion of affected communities*
- *More effective advocacy*
- *Joint strategic and operational planning*



Source: HumanitarianResponse.info

Figure 1: The cluster system

As part of the humanitarian reform process, the *Global WASH Cluster (GWC)* was formed in 2006 upon an existing WASH sector working group (GWC, 2011, p.10). *The United Nations Children's Fund (UNICEF)* is acting as its lead agency. On a country level, the lead agency of the WASH cluster may vary and be also a government authority or a major (international) non-governmental organisation ((I)NGO). After years of evolution and reaching a certain point of maturity, in 2011 and for the first time, the GWC (including its 26 active members) has adopted a strategic plan (GWC, 2011) covering the period until

2015. According to it, the GWC's vision is to *"ensure that all those affected by humanitarian crises have equitable access to timely and comprehensive WASH services"*, reflecting *"the collective cluster agencies' commitment to support progressive realisation of the right to water, sanitation and hygiene in emergencies, and the importance of WASH in facilitating physical protection, social stability and psychological recovery in the midst of crisis"* (GWC, 2011, p.15). Accountability, learning and evaluation (as central themes of the thesis) constitute one out of five priority areas of work. The respective objective is formulated as:

"Accountability within and between WASH cluster agencies is enhanced and facilitated by the lessons learned and best practice captured through structured and formal evaluation and documentation of major responses" (GWC, 2011, p.3).

The way WASH clusters operate is continuously evolving and in itself a big learning process informed by every new deployment. Although there is a general perception that the positive effects of clusters outnumber the negative ones, the still top-down orientation of the humanitarian system and the risk of undermining local capacities are a repeated concern (Harvey, et al., 2010, pp.10-11). The latter is also voiced in some of the researched evaluation reports. E.g. Van der Wijk, Mwezi and Kazan (2010, p.24) find that *"coordination and communication appears stronger on the community level than at the level of the clusters […] technical staff takes part in both the WASH and Health Clusters, but they expressed the feeling that the clusters are often dominated by the International agencies and INGOs and that local NGOs are left out of major discussions and decisions. None of the INGOs active in the region have taken the initiative to harmonize and coordinate their projects with those of local NGOs"*.

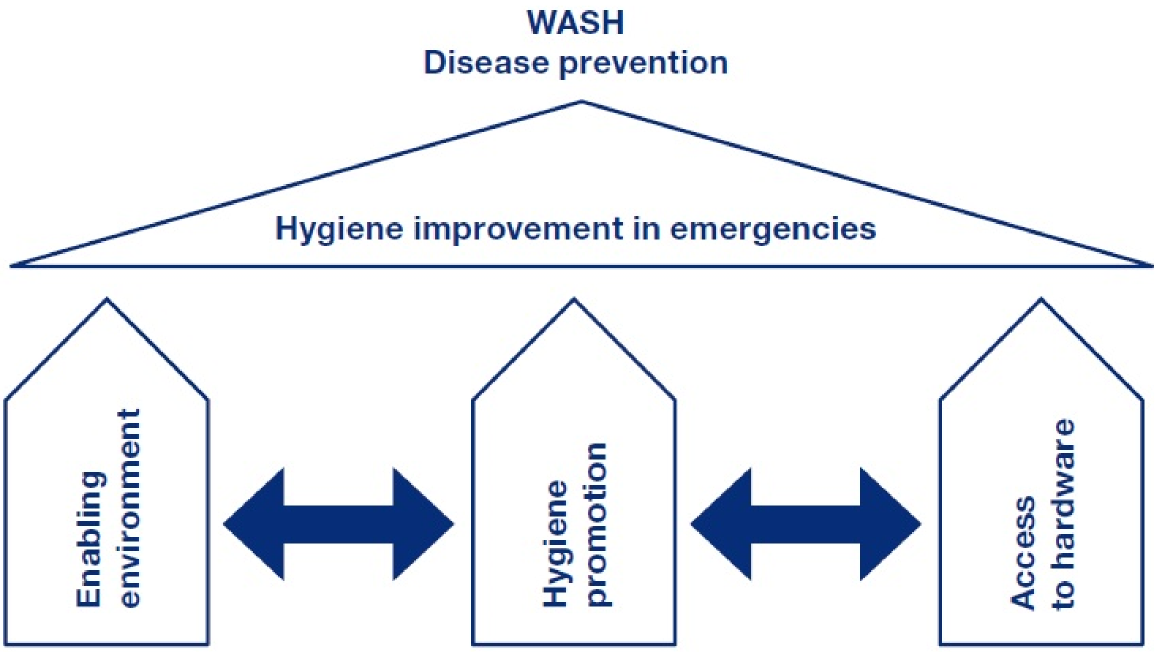
Such paradoxes – "that humanitarian assistance undermines rather than strengthens local capacity" (Stokke, 2007, p.16) – do not just affect overall performance of all interventions but have momentous adverse consequences from an LRRD perspective. In general, inter-agency and inter-cluster or inter-sector coordination remains an area with a lot of room for improvement. For WASH, especially *Health, Education, Shelter and Camp Coordination/Management* are main clusters or sectors (if there are no clusters) to coordinate with. The GWC also recognises its advocacy role for improving national and subnational WASH coordination platforms and the need to "go beyond cluster coordination and look at sector coordination" (GWC, 2013, p.5). As this institutional environment for WASH in emergencies has gradually developed, so have the objectives of humanitarian response which have been predominantly hardware centred and quantitative not so long ago. With the latest edition of the *Humanitarian Charter and Minimum Standards in Humanitarian Response* the main objective of WASH emergency response can be said to be (The Sphere Project, 2011, p.84):

To reduce the transmission of faeco-oral diseases and exposure to disease-bearing vectors through the promotion of:

- *good hygiene practices*
- *the provision of safe drinking water*
- *the reduction of environmental health risks*

- *the conditions that allow people to live with good health, dignity, comfort and security.*

Depending on the type of an emergency, which can broadly categorised in either being acute or chronic/protracted, WASH interventions typically employ different approaches. They do however have in common the same aim: Promoting good personal and environmental hygiene in order to protect health (The Sphere Project, 2011, p.88). *Figure 2* visualises the overall goal and main components of the intervention logic of any WASH programme which is concerned with *hygiene promotion (HP), water supply, excreta disposal, vector control, solid waste management and drainage*. In brief, realising hygiene improvement in emergencies leading to WASH disease prevention requires a combination of the three interdependent pillars: Access to hardware (sanitation facilities, water supply systems, etc.), hygiene promotion (community mobilisation, information, education and communication, etc.) and an enabling environment (financing, institutional strengthening of WASH, etc.).



Source: The Sphere Project, 2011, p.88

Figure 2: Goal and intervention logic of WASH programmes

This intervention logic is based on the *Hygiene Improvement Framework* (EHP, 2004), developed a decade ago by major UN and affiliated institutions active in WASH and the U.S. Agency for International Development. It is a clear expression of the integral role that hygiene promotion plays in a successful WASH intervention. UNICEF, one of the contributing UN organisations with long-standing experience in emergency response dating back to the late 1960s, reasons: “[T]he single most important lesson learned throughout the world” is “that water and sanitation facilities on their own do not automatically result in improved health” (EHP, 2004, p.v). Nevertheless, recent practice (as depicted in the evaluation reports) suggest that much works remains to be done to fully incorporate this lesson learned. Hygiene promotion still tends to get underfunded and considered rather as an add-on to emergency water supply and sanitation (see e.g. Barham, Nabunny and Philpott, 2011a, p.19 or Grayel and Mattson, 2012, p.7). Therefore, it is not surprising to see elementary hygiene promotion issues frequently raised in a global WASH gap analysis conducted in 2013 (*Table 1*).

Ranking	Issue raised	Frequency
1	Latrines where no pits are possible (urban, high water table, floods, rock, snow, sands)	37
2	Community participation/empowerment of vulnerable groups including monitoring and evaluation (M&E) from beginning	27
3	Latrine emptying/desludging	26
4	HP - importance of context, understanding, including socio-anthropology	26
5	Community-Led Total Sanitation (CLTS) and sanitation marketing	23
6	Urban alternatives for excreta disposal	22
7	Exit strategies and sustainability issues to be considered from the start of a response	22
8	Final disposal options after desludging + treatment	19
9	Further development of non-toilet options/early response/mobile options	19
10	Hand washing hardware + promotion & sustainability + soap/non soap options	18
11	Water treatment - bulk vs. point-of-use (POU), filters, household water treatment (HHWT), cost, sustainability, mobile unit	18
12	Need low-tech solutions acceptable and sustainable by locals	18
13	Emergency-development continuum including listening to existing field knowledge	18
14	Ecological sanitation (EcoSan) + biogas – eco-friendly solutions	17
15	Latrines - facilitating anal cleansing	16
16	Water management including to involve private sector	15
17	General drainage from showers and wash units	15
18	Menstrual hygiene provision	14
19	Maintenance, (latrines) sustainability, cleaning, cash for work	13
20	Maintenance of water sources and supply, spare pump parts, monitoring	12
21	Community behaviour change	12
22	Environmental concerns (wastage at pump, poor drainage)	12
23	Shared & family latrine sustainability/replicability	10
24	Solid (rubbish) waste management systems including possible recycling	10
25	HP extended to schools and community groups, health clubs	10
26	Improved integration of disaster risk reduction (DRR) in WASH and enable community water safety plans	10

Data source: Bastable and Russell, 2013, p.14

Table 1: Global gap analysis in emergency water, sanitation and hygiene promotion

The aim of this analysis was "to identify the major challenges that require innovative solutions in the Humanitarian WASH sector" (Bastable and Russell, 2013, p.1). The data gathered from beneficiaries through focus group discussions, literature review and surveys at (I)NGOs, donors and WASH clusters are presented as categorised issues in *Table 1*. It shows that sanitation issues are identified as the major area with gaps and potential for innovation, followed by HP issues. The gap analysis points out a lack of excreta disposal solutions in difficult environments and the general need for evidence on which approach or technology works best in a particular context. Accountability related issues, such as community participation or monitoring and evaluation (M&E), and LRRD related issues, such as exit strategies and sustainable technologies, form further prominent themes. According to a GWC workshop conducted as part of the gap analysis project, the important issues for which to seek innovative solutions for scaling up and sharing with the whole sector are (GWC, 2013, p.6):

- *Excreta disposal in difficult environments including final deposition site for desludged excreta*
- *WASH in urban environments*
- *Transition from emergency to development – policy and practice and sustainability*
- *Evidence base for what works in hygiene promotion and community mobilization*
- *Capacity building of local partners/national training initiatives*
- *Hand washing – promotion of no soap and no water options*
- *Development of exit strategies from WASH and Cluster*
- *Management of facilities during and after*
- *WASH response to drought*
- *Targeting gender and vulnerable groups in emergencies*

2.2 Evaluation of Humanitarian Action: Issues to Consider and the Application to WASH Emergency Interventions

To continue, the global gap analysis in emergency WASH identified further issues beyond its scope. Among them, evaluation and engagement with academia to help building evidence is noted. The reason is straightforward: "Any innovation requires robust evaluation." (Bastable and Russell, 2013, p.26). In any case, generating evidence through evaluations is confronted with several constraints worth bearing in mind. First of all, for EHA, the emergency context exacerbated by poor evaluation planning brings along a wide range of common challenges. To name a few, e.g. evaluation suspension and restricted access to field locations due to a volatile security situation, too short time frames, quality, inaccessibility or lack of documentation and data or logistic constraints are mentioned in the researched evaluation reports. Secondly, demonstrating evidence raises difficult methodological problems which are also linked to data availability as noted above. Despite improvements in the quality of evaluations, recent discussions (e.g. Guerrero, Woodhead and Hounjet, 2013 or Darcy and Knox Clarke, 2013) suggest that there is yet more than a grain of truth in Lindahl's (1998, p.3) statement from the early days of EHA: "Evaluations tend to be an *art* rather than a *science*, and while good *art* can provide valuable and sometimes highly revealing impressions, it is still coloured by the temperament of the artist.".

To properly grasp the term evidence, there can be only evidence *for* something. It is "information or analysis that goes to support a particular proposition or claim" (Darcy and Knox Clarke, 2013, p.7). According to Darcy, et al. (2013, p.19) three main proposition types can be distinguished in the humanitarian sphere:

- *That an actual imminent crisis exists;*
- *That a given form of response will be (or has been) effective in preventing or mitigating the worst aspects of this crisis;*

- *That a given form of response is the most appropriate in context, in view of effectiveness plus other factors: alternative response options, local preferences and responses, feasibility, standards, cost, etc.*

For this work, the second and third proposition are relevant. The first proposition, i.e. evidence that a situation is critical for those affected is of minor concern – given the fundamental importance of WASH (not just during crisis), the preventive character of WASH interventions for health and that WASH emergency response rather faces the problem of targeting because of low pre-crisis coverage of WASH infrastructure and services. The central conclusion to these propositions is that they require different kinds of evidence to substantiate them which uncover weaknesses and gaps of humanitarian action and its evaluation practices (Darcy and Knox Clarke, 2013, p.38).

Firstly, the “preferences and attitudes of crisis-affected people must be factored into analysis” (ibid.). At the same time it is well documented that humanitarian organisations continue to fail in consulting them or to use their input in programming (Taylor, et al., 2013, p.10). The findings from the analysed evaluation reports in this regard are mixed but tend to support this view. Secondly, “mixed methods of enquiry (qualitative and quantitative) will almost always be required to gain a true picture of what is happening” (Darcy and Knox Clarke, 2013, p.38). However, as Darcy and Knox Clarke (2013, p.39) point out, EHA tends to rely “almost exclusively” on qualitative methods. The methodologies of the evaluation reports illustrate the same. Finally, researchers emphasise to avoid generating evidence within a sector or particular area in isolation from others (Darcy and Knox Clarke, 2013, p.38). For WASH this means that at least health and livelihoods as well as cross-cutting issues such as gender, disaster risk reduction (DRR) or environment should be considered in any attempt to establish evidence.

According to its purpose, two key types of evaluations can be distinguished in the humanitarian sphere (Buchanan-Smith and Cosgrave, 2013, p.22):

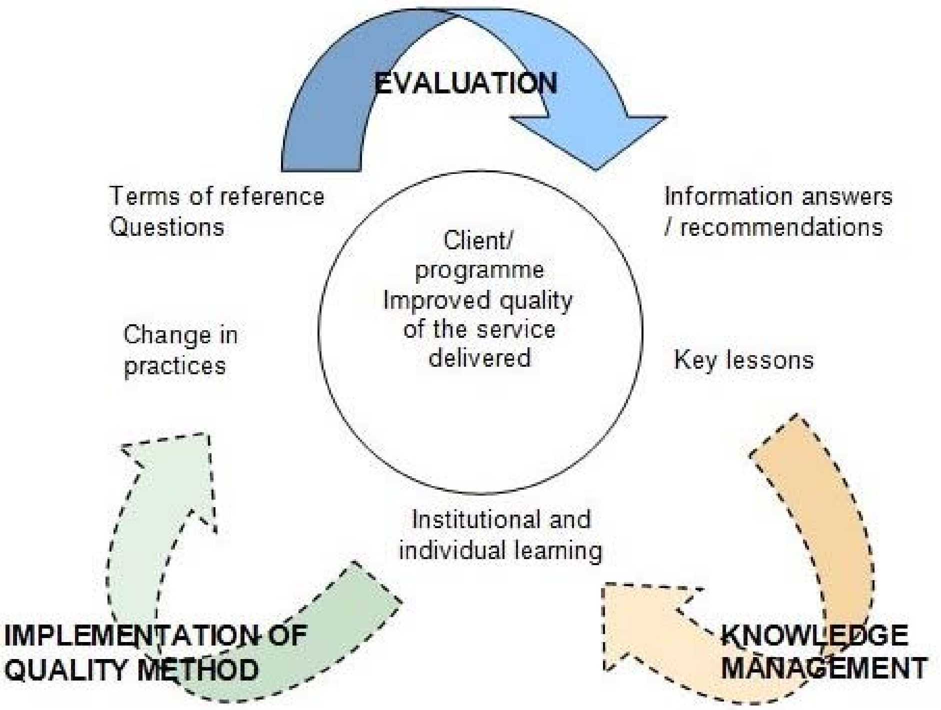
- ***Accountability-oriented evaluation***

  An evaluation of how well resources have been applied (also called a summative evaluation).

- ***Learning-oriented evaluation***

  An evaluation designed to facilitate individual, group, or organisational learning (also called a formative evaluation).

Although in practice accountability and learning are difficult to separate (ibid.), evaluation research keeps finding that accountability and learning aims are rather conflicting each other and trying to cover both in one evaluation will only contribute to poor utilisation (Borton and Robertson, 2002, pp.182-184; Hallam, 2011, pp.10-18). In several evaluation reports the purpose of the evaluation remains unclear and there appears to be a tendency to orient evaluations demanded for accountability reasons towards learning. A far more important insight from EHA research is that “[i]f we continue to expect evaluation to cover most of the accountability and learning needs of the sector, we will be disappointed.” (Sandison, 2006, p.139). The extent to which accountability-focused evaluations in its common form as donor induced single-agency evaluations are truly serving their purpose is questionable. Besides regarding the evaluation quality to be expected, particularly in a sector such as WASH where gaps and learning needs are big, not focussing an evaluation on learning might be a bad investment as well as missing a rare opportunity. For example ECHO – one of the biggest humanitarian donors – emphasises as one of the first sentences in the preamble to its guide *Evaluation of Humanitarian Aid by and for NGOs* (DG ECHO, 2007): "An evaluation is first and foremost an opportunity to **LEARN**!".



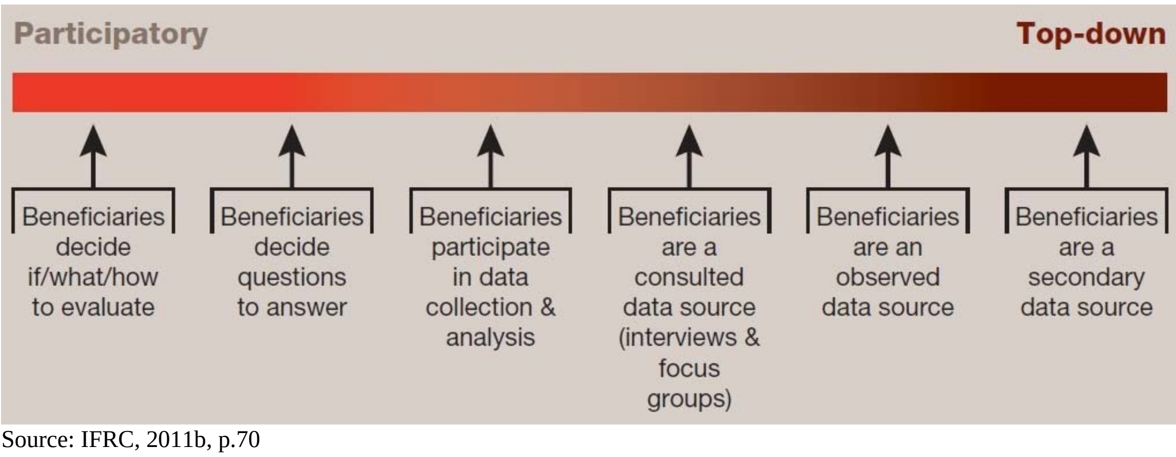
Source: Maury, 2009, p.12

Figure 3: Theoretical diagram of the learning process, from the evaluation to a better quality of service delivered – benefits and "missing links"

However, placing evaluation as a major tool for learning does not automatically lead to learning and changes (see the theoretical process from evaluation to change in practices in *Figure 3*). This lesson is known at least since the late 1990s. Van Brabant (1997, p.31) identifies two problems for learning from evaluations: "[E]valautions, particularly in the humanitarian field, seem to suffer from methodological anarchy" and "even where an evaluation can be methodologically convincing, poor monitoring and recording systems may deny it the relevant information". While the latter is an on-going concern, recent evaluation research (e.g. Hallam and Bonino, 2013) does not suggest "methodological anarchy" – which appears to have been a specific problem in the early days of EHA – as a root cause for non-learning. Producing "high-quality evaluation products" is identified as "only half of the challenge" (Hallam and Bonino, 2013, pp.11-12). The other half is about "strengthening the capacities of individuals, teams and organisations to plan, commission, conduct, communicate and follow-up credible and timely evaluations" (ibid.).

Several of these aspects are framed by Maury (2009) as *Knowledge Management* and the *Implementation of a Quality Method* – the major missing links in the learning process from evaluations to improved performance (*Figure 3*). The author argues that the humanitarian sector can yet much learn from business where the formalisation/structuring and sharing of information, knowledge and experience, etc. and the implementation of quality assurance methods are common sense because they are essential to "survival" (Maury, 2009, pp.12-13). Maury (ibid.) concludes that if nothing is done in terms of knowledge and quality management, "the hopes of progress that evaluations gave rise to will be dashed for a long time to come". Concerning WASH, the past years have shown increased attention towards these missing links. The GWC's strategic plan (GWC, 2011) identifies knowledge management as one of its core initiatives and to this effect field level learning papers[1] have been published. So, there is still hope for this research to be relevant whose focus and scope is located before the beginning of the learning process – marked in *Figure* 3 with the formulation of the terms of reference and questions for the evaluation. It is believed that evaluations should provide a source for learning.

Next to the WASH learning needs mentioned in the previous section, a general but particularly important issue for WASH is learning from the affected population. A study conducted ten years ago revealed though interviews that humanitarians – while stressing participation (at least in policy) – don't even consider learning from those they aim to assist (Beck and Borton, 2004, p.53). How exactly evaluation can be utilised to provide learning from crisis-affected people is up for discussion, but there is common agreement that they should be at the centre of the evaluation process. According to Hallam and Bonino (2013, p.62), only modest improvements have been achieved in this regard over the years. Those affected by crisis are usually only involved as informants rather than in setting the evaluation agenda (ibid.). The participatory continuum in evaluations is shown in *Figure 4*.



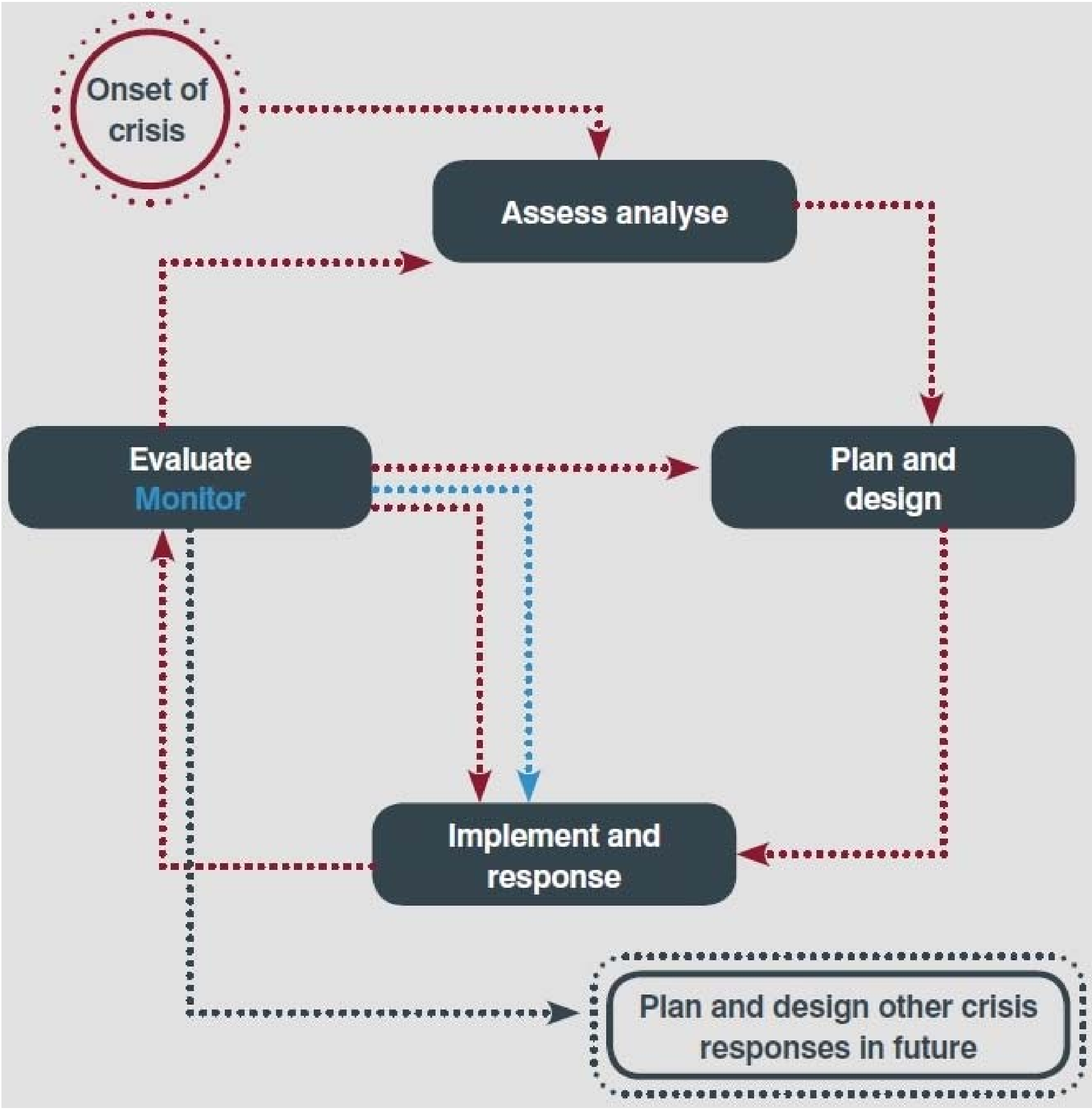
Source: IFRC, 2011b, p.70

Figure 4: The participatory continuum

The researched evaluations echo above findings. Usually or in almost all cases, affected people have been interviewed or were participating in focus group discussions with

[1] One of the most recent examples of such publications is *Knowledge Management in Practice. Implementing Effective Knowledge Management in Emergencies: A Case Study from Somalia WASH Cluster* (Harries, 2013).

evaluators, i.e. being not more than informants. The potential advantages and disadvantages of participatory evaluations or M&E altogether are well understood and discussed in agency's own M&E guides (see e.g. IFRC, 2011, pp.69-71). And of course, the degree of participation will always depend on the emergency context. However, it seems to be a reasonable line of thought, that not achieving higher degrees of participation of those affected by disasters in evaluations does have to do with an organisation's timely planning. This includes making the necessary resources available too. The WASH evaluation reports cover interventions with durations ranging between nine months and more than two years, and yet in some instances poorly timed or planned preparations for the evaluation are noted.

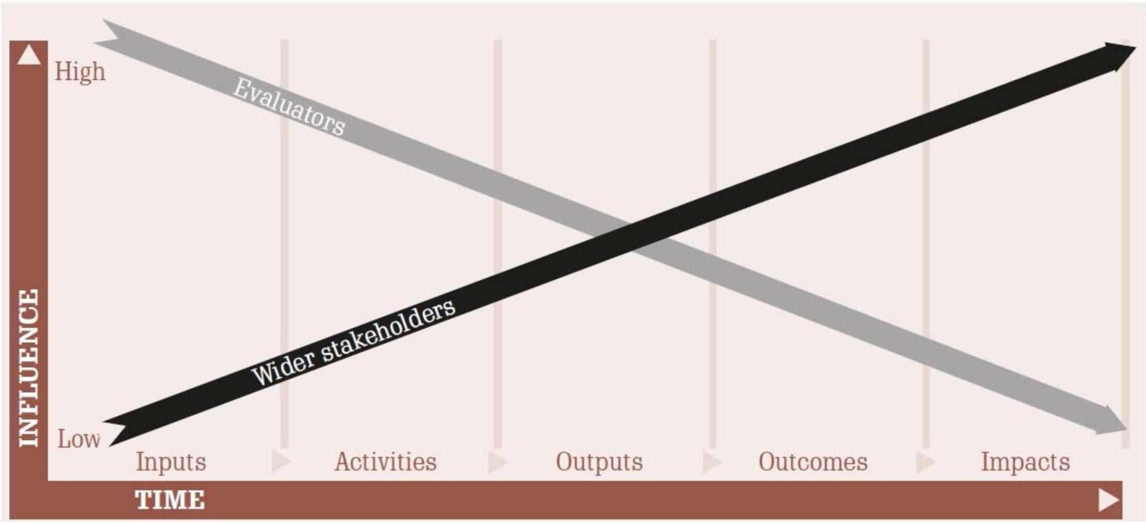


Source: Buchanan-Smith and Cosgrave, 2013, p.22

Figure 5: Monitoring and evaluation in the emergency response cycle

The idealised role of M&E in the emergency response cycle is illustrated in *Figure 5*. From a chronological point of view, the first evaluation taking place after the onset of a disaster is an evaluation of the needs, which however is commonly referred to as assessment. Needs assessments are typically based on indicators and not on criteria such as suggested by the OECD-DAC. Though in practice, as Darcy and Knox Clarke (2013, p.19) explain, “assessments often combine situational analysis with response analysis”, and the “relative feasibility, appropriateness and cost-effectiveness” of different intervention options may form part of it. But the evaluations of concern in the following work are those carried out either during emergency response to inform intervention design and implementation based on needs and context analysis, or at the end/after an emergency project to inform future response (*Figure 5*). In theory, if effective monitoring systems are implemented from the very beginning of emergency response, such evaluations might not be necessary in the first place as they are unlikely to add additional value. Anyhow, humanitarian practice is quite different. Poor or even lack of monitoring are a continuous theme throughout the researched evaluations. Actually, sometimes it seems that evaluations are the only way to establish a sense of what happened, why it happened and how it affected performance. The evaluation of a cholera response from 2008 to 2010 in Zimbabwe conducted by DeVillez, Bousquet and Nyagwambo (2011) provides a good example of this.

Fundamentally different from the evaluation approach described in *Figure 5* are strategically selected evaluations (Buchanan-Smith and Cosgrave, 2013, pp.22-23). Buchanan-Smith and Cosgrave (ibid.) note that the Swedish International Development Cooperation Agency has successfully experimented with this approach, which starts with determining what operational units would like to know and how evaluation could help. Strategic selection of evaluations could as well be a promising approach to tackle several of the WASH gaps identified by Bastable and Russell (2013) and the GWC (2013). In the end, whatever approach is applied, “the evaluation is but a catalyst, influencing the beginning of a process and increasingly ceding control to the impact of other actors and other forces“ (Sandison, 2006, p.136). *Figure 6* illustrates the decreasing influence of evaluators in the change process.



Source: Sandison, 2006, p.136

Figure 6: The decreasing influence of evaluators in the change process

To make better use of the potential influence of EHA in decision making processes and improving performance when still achievable, evaluations in the early phase of emergency response are increasingly executed. Over the last three years, RTEs have been pushed by e.g. the *Inter-Agency Standing Committee (IASC)* – the primary mechanism for inter-agency coordination of humanitarian assistance for key UN and non-UN partners – and are further elaborated (Darcy and Knox Clarke, 2013, p.31). An RTE is understood as “an evaluation in which the primary objective is to provide feedback in a participatory way (i.e. during the evaluation fieldwork) to those executing and managing the humanitarian response” (Cosgrave, Ramalingam and Beck, 2009, p.10). *Figure 7* summarises the logic behind RTEs which can be differentiated in forward- and backward-looking aspects. The first pilot guide for RTE in humanitarian action published by ALNAP outlines three key benefits of this evaluation type (Cosgrave, Ramalingam and Beck, 2009, p.14):

- ***Timeliness***

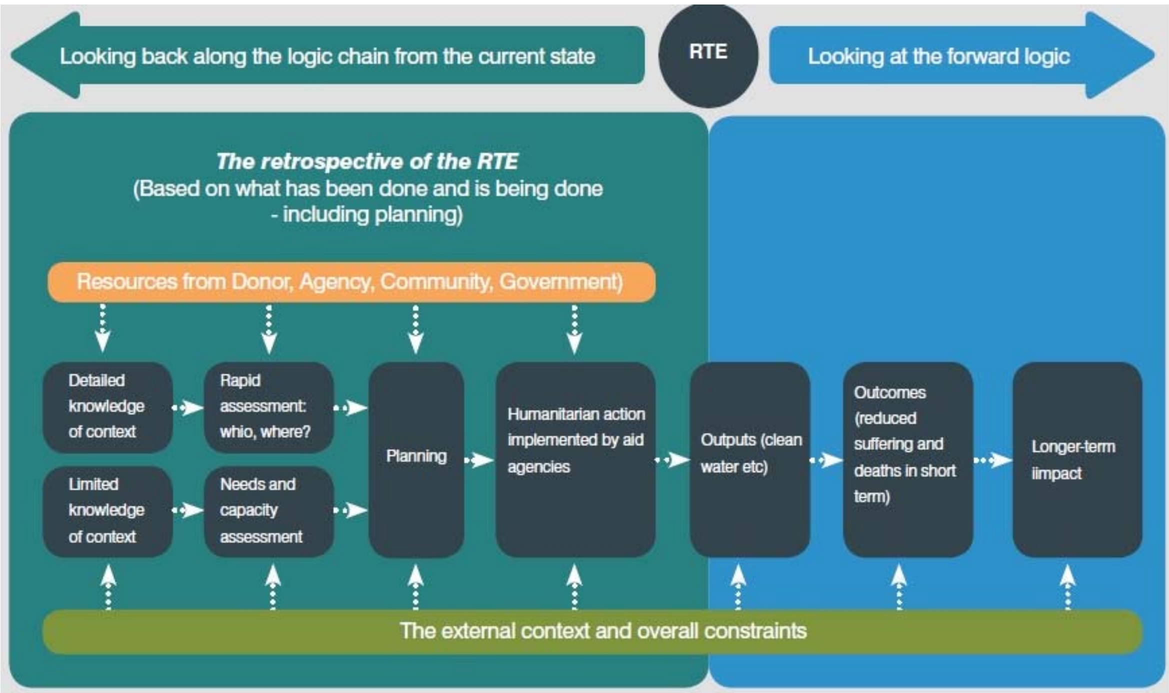
  RTEs bring in an external perspective, analytical capacity and knowledge at a key point in a response.

- ***Perspective***

  RTEs reduce the risk that early operational choices bring about critical problems in the longer term.

- ***Interactivity***

  RTEs enable programming to be influenced as it happens, allowing agencies to make key changes at an intermediate point in programming.



Source: Buchanan-Smith and Cosgrave, 2013, p.184

Figure 7: Forward- and backward-looking aspects of real-time evaluations

Currently, RTEs are mainly applied in responses to major sudden onset disasters respectively are only required for such by the IASC and agencies' evaluation policies (see e.g. IFRC, 2011a, p.10). They are a relatively new form of evaluation in the WASH sector. The two RTEs included in the selected evaluation reports have been conducted in 2010 and 2011 in the aftermath of the earthquake in Haiti. The final part of this section briefly summarises the results of applying the *ALNAP Quality Proforma* (*Annex 1*) to them and the eight ex-post evaluations. In relation to the following five sections of the Quality Proforma, the respective requirements (if applicable) that were poorly or unsatisfactorily covered in the reports are listed with their frequency in brackets:

- **Section 1: Assessing the Terms of Reference (ToR)**
  - 1.1 The Terms of Reference (4/10)
  - 1.2 Expectation of good evaluation practice (4/10)
- **Section 2: Assessing Evaluation Methods, Practice and Constraints**
  - 2.1 Nature, make up and appropriateness and biases of the evaluation team (7/10)
  - 2.3 Appropriateness of the overall evaluation methods (10/10)
  - 2.6 Evaluation constraints (3/10)
- **Section 3: Assessing Contextual Analysis**
  - 3.2 Past involvement of the agency and its local partners (2/10)
- **Section 4: Assessing the Intervention**
  - 4.1.i The agency's guiding policies and principles (1/10)
  - 4.1.ii The agency's management and human resources (3/10)
  - 4.2.i The needs and livelihoods assessments that informed the intervention (3/10)
  - 4.2.ii Intervention objectives (1/10)
  - 4.2.iii Programme cycle processes (3/10)
- **Section 5: Assessing the Report**
  - 5.1.i Secondary sources (5/10)
  - 5.1.ii Conclusions (2/10)

The subsections *4.3 Application of EHA Criteria* and *4.4 Consideration given to Cross-cutting Issues* are not included on purpose. These are dealt with in detail in *Chapter 3* and *Chapter 4*. Moreover, in this regard the Quality Proforma is not up-to-date, i.e. it does not reflect the suggested use of EHA criteria according to lessons learned (e.g. Buchanan-Smith and Cosgrave, 2013). In general, it is difficult to assess the quality of an evaluation report with a list of evaluation criteria and cross-cutting themes to be covered, mainly because it is often not possible to determine the applicability of a criterion or specific theme in the first place. Depending on the context, some of the OECD-DAC criteria and maybe also some of the cross-cutting themes are rather nice to have included in an evaluation than a must have considering overall evaluation constraints. Specifically for RTEs, the Quality Proforma matrix is not fully applicable and should be adapted to suit them in better ways if this tool is to be used to assess their quality.

The list of weak spots eminently highlights that for evaluation reports it is still not common to

- at least have the terms of reference attached (which definitely helps to get a better picture of the evaluation and its purpose, and of which not all of the intended audience might be aware),
- critically reflect on the nature, make up and appropriateness and biases of the evaluation team,
- consider the appropriateness of the overall evaluation methods,
- and use secondary sources to support findings, conclusions and recommendations.

Above findings reveal that there remains considerable room for improvement in the quality of evaluation reports, yet the main product of an evaluation. An EHA course reference manual summarises: "A good report reflects well on the evaluation manager and the evaluation team, a bad report does otherwise." (Buchanan-Smith and Cosgrave, 2010?, p.128). The listed weaknesses are avoidable and needlessly affect the credibility of evidence. Moreover, they add up to the methodological constraints in generating the best available evidence base for learning through evaluations in data-poor and operationally fluid environments. But regarding evaluation utilisation, this is just one out of many issues. For all evaluations, first and foremost learning-oriented evaluations, reports should not be the only output. The expected evaluation outputs – thus far mentioned in the annexes of the evaluation reports – suggest, that the submission of the final evaluation report marks the end of the evaluation exercise, often only accompanied by formal dissemination activities such as stakeholder meetings. These are undoubtedly less adequate ways of harnessing the knowledge created through evaluations. To overcome such practices, e.g. Hallam and Bonino (2013, pp.67-72) have compiled dozens of pointers for disseminating evaluation findings effectively.

2.3 The OECD-DAC Reference Framework and Criteria for Evaluating Humanitarian Action

With Buchanan-Smith and Cosgrave (2013, p. 52), two fundamental evaluation criteria can be distinguished: *Quality* and *value*. The OECD-DAC criteria for EHA originate[2] from the evaluation principles set out for the DAC members in 1991 which reflected the most common problems in development assistance. These were later refined into the four quality criteria *relevance, effectiveness, sustainability and impact*, and the value criterion *efficiency* (ibid.). Subsequently, the five criteria for development evaluations have been adapted for

[2] For the ones interested in the processes, institutions and initiatives involved in the long accruement of the OECD-DAC criteria, e.g. the preface to the Good Practice Review *Evaluating Humanitarian Assistance Programmes in Complex Emergencies* (Hallam, 1998, pp.5-6) alludes some of these.

the evaluation of complex emergencies (Beck, 2006, p.10), becoming the set of seven criteria (*relevance/appropriateness, connectedness, coherence, coverage, efficiency, effectiveness and impact*) which is examined in the research. These evaluation criteria for humanitarian interventions – henceforth called the OECD-DAC criteria – guide *Chapter 3* which includes their definitions.

Since the first EHA guidelines and good practice notes were published in the late 1990s (e.g. Hallam, 1998), the quality of evaluations was an on-going concern and has been pointed out by ALNAP's annual meta-evaluation reviews (see e.g. Borton and Robertson, 2002, pp.161-190). Hence, further materials on the proper use of the OECD-DAC criteria were elaborated. In fact, these would suggest to better speak of the OECD-DAC framework than just the criteria. The OECD-DAC framework underlying the meta-analysis is the one drawn up by Beck (2006), which has become a standard guide for EHA and builds on previous influential publications. Except for minor issues, such as covering specific themes under a different heading, it is identical to the ALNAP Quality Proforma published a year earlier, and which is also referenced by Beck as a checklist of good practice for EHA.

The rationale of the OECD-DAC criteria is based on the idea of using the criteria in combination. As Beck (2006, p.18) explains: "The DAC criteria are designed to promote comprehensive evaluation of humanitarian action. For this reason, the criteria are complementary. […] Using the DAC criteria in combination will help to ensure that an evaluation covers all areas of the intervention.". It is important to keep in mind that this rationale is rooted in the mindset of the *Logical Framework Approach (LFA)* respectively *Project Cycle Management (PCM)* based on the LFA. These project design and management tools have been established and widely-used in the development sector well before the emergence of the OECD-DAC criteria. This simply means that the latter have been conceived at the backdrop of evaluating a detailed operational plan, based on the "bottom-up chain of effects" in classical LFA (inputs → activities → outputs → outcomes → impact) and its corresponding intervention logic, in a comprehensive and complementary manner (Eggers, 2009, pp.117-118).

The critical examination of the OECD-DAC criteria by researchers in the relief to development contiguum can be broadly separated into two strands. The first strand relates to confusions about specific terms, definitions and guidance for their application, leading to misunderstandings in the use of the evaluation criteria. The second strand is more fundamental and addresses the comprehensiveness of the OECD-DAC framework in terms of covering quality aspects in evaluations. The main points of both are outlined below:

- **Terminology, definitions and use**

  There is a general observation that several of the OECD-DAC criteria are not well understood, and that they are often used mechanically (i.e. the whole set of criteria is applied regardless of the context) rather than intelligently (i.e. using only the relevant criteria in more creative ways) by both, evaluators and evaluation managers (Borton and Robertson, 2002, pp.182-183; Beck, 2006, pp.10-19; DG ECHO, 2007, p.51). Borton and Robertson (2002, p.182) assert that the "criteria as adapted invite a focus on determining 'what happened' rather than 'why it happened' […] reinforcing the lack of attention given to process" and that this is "to some extent encouraged by the nature of the DAC criteria definitions". Beck's (2006) OECD-DAC framework targets these concerns explicitly.

De Geoffroy and Kauffmann from *Groupe Urgence, Réhabilitation and Développement (URD)* (2009, p.15) bring up another perspective in stating: "In particular, the fact that the criteria are represented by single words often leads to misunderstanding and misinterpretation regarding their meaning. Therefore, the DAC criteria are perceived by the majority of humanitarian actors as only being useful to evaluators who are experienced and highly-trained. This runs counter to the idea that a quality framework is a set of common values which are widely shared between stakeholders.". Out of this observation, the Groupe URD developed the *Quality COMPAS* as a project management and evaluation method (see the Group URD website under *http://www.urd.org* for resources on the Quality COMPAS). The approach is based on twelve criteria which take the form of short sentences (expressed as requirements) instead of using a single word with a corresponding definition. These criteria are assumed to tackle the important issues in humanitarian action. For each criterion a set of key processes is assigned, also formulated as requirements. Despite this arguable approach and the fact that the OECD-DAC framework is dominant within EHA and humanitarian agencies, the underlying idea of the Quality COMPAS method to have easier to handle evaluation criteria is a valid point, particularly for an attempt to involve crisis-affected communities in setting the evaluation agenda.

- **Lack of covering the quality of process**

  The reason for the development of the Quality COMPAS by Groupe URD was precisely the lack of a general quality reference framework for humanitarian action that covers not only impact and results, but also structure and processes (De Geoffroy and Kauffmann, 2009, p.14). De Geoffroy and Kauffmann (2009, p.15) argue that the OECD-DAC framework has been useful, but would need to be revised to make it a "genuine [quality] management tool for organisations". Similarly, Chianca (2008, pp.47-48) identifies the *quality of process* as a missing key criterion, although the author refers to the established set of five OECD-DAC criteria for development projects. The author (ibid.) proposes a set of quality aspects to be included in such a criterion and acknowledges that the OECD-DAC framework elaborated by Beck (2006) includes the quality of process by incorporating cross-cutting themes to be covered in evaluations.

Chianca's (2008, pp.44-48) other proposals (e.g. focus on needs or including comparisons with alternatives within the criterion efficiency) are more or less also included in Beck's evaluation framework. There is just one interesting exception. In addition to the quality of process, Chianca (2008, p.48) proposes another new criterion (termed *exportability*) to target the extent to which an intervention as a whole or some its elements are transferable to other settings. While this is a noble idea, as outlined above, the main challenge for EHA and the humanitarian community in general is still a different one, i.e. finding out what works best in different settings first.

Eggers (2009), who commented on Chianca's proposals, points out that the OECD-DAC criteria are to be viewed on the basis of project planning tools such as LFA and PCM, the other side of the coin. As e.g. quality aspects of interventions are already supposed to be included in the planning documents, there is no need to revise the OECD-DAC criteria (Eggers, 2009, pp.117-120). Hence, as emergency response is not based on such elaborated planning, to cover the quality of process in EHA, a range of cross-cutting themes as well as pointers for good practice, such as assessing the intervention against humanitarian standards, need to be perceived together with the evaluation criteria (see Beck, 2006).

For the sake of completeness, it should be mentioned that project planning tools based on the LFA are widely used in the humanitarian sector, but they serve rather the purpose of providing an overview of the response activities (Ramalingam, et al., 2009, pp.55-56). As the evaluation reports and a recent survey (Guerrero, Woodhead and Hounjet, 2013) demonstrate, they are inappropriate for developing a meaningful logic from inputs to impacts which also works in volatile emergency contexts where decision with high potential impacts often have to be made early and quickly. And in general, according to Guerrero, Woodhead and Hounjet (2013, pp.9-10), a framework that helps to measure and reward quality is missing. In relation to M&E the authors (ibid.) state: "Humanitarian staff know that the indicators being measured say little about the quality of programmes, and that quality seldom influences the ability of organisations to continue their operations, because funding is not conditioned by what these indicators say.".

For the reasons explained above and because evaluations are more accountability driven than monitoring, EHA constitutes the only strong tool to assess the quality of emergency interventions. This use however is compromised by weak monitoring and missing incentives. At the same time, the widely-used OECD-DAC evaluation framework is a rather intricate web of interrelated criteria, cross-cutting themes, standards and best practice guidance which many find difficult to navigate through. As humanitarian action is a constantly evolving field and new challenges emerge while old ones persist, further reviews of the OECD-DAC criteria are to be expected. E.g. Cosgrave, Ramalingam and Beck (2009, p.62) sum up: "If the humanitarian action evaluation guidelines were rewritten today, they would doubtless include accountability as a criterion". But multiplying criteria is "hardly helpful to simplifying the evaluation process" (DG ECHO, 2007, p.51). ECHO's EHA guide argues in favour of a limited set of generic criteria as convenient entry points into evaluation and acknowledges that in parallel a much larger number of "sub-criteria", adapted to various situations, could be useful (ibid.). The proposed generic criteria – ECHO suggests they might be called *Humanitarian Aid Criteria for Evaluation (HACE)* – are (DG ECHO, 2007, pp.52-53):

- **For outcome evaluations:**
    - *Effectiveness* to outline the particular importance of results (a generic alternative could be *Impact*)
    - *Relevance* to comprise every effort to identify needs and priorities, relevant local or wider contexts and their implications, adequate entry and exit strategies, coverage and coordination, etc. (an alternative generic term could be *Strategy*)

    - *Values* to address and enforce humanitarian principles, laws, standards and codes
- **For process evaluations:**
    - *Efficiency* (including the quality of process) to encompass the quality of the internal organisation of an agency, its capacity to deliver outputs as expected and its adequacy to fulfilling the demanding needs of conducting humanitarian operations

As recent EHA guides (Buchanan-Smith and Cosgrave, 2013; Cosgrave, Ramalingam and Beck, 2009) emphasise, the OECD-DAC criteria offer just one possible framework for evaluation. Other possible frameworks can be distinguished in (Buchanan-Smith and Cosgrave, 2013, pp.59-60):

- Broad normative frameworks (e.g. humanitarian principles)
- Broad conceptual frameworks (i.e. a model of how a system is thought to work)
- System-wide standards and guidelines (e.g. HAP and Sphere standards)
- Sectoral standards (e.g. Sphere minimum standards in WASH)
- Agency standards and guides

EHA guides also consider these frameworks to be used in addition to the OECD-DAC criteria, or as a substitute if appropriate. For this research, the combination of the OECD-DAC criteria with quality criteria (standards) is important to address and discuss the lack of covering the quality of process as outlined above. This combination is actually already suggested through the OECD-DAC framework of Beck (2006), although probably not always understood like that. The most recent evaluation reports indicate, that the Sphere and HAP standards have become more commonly used in evaluations as an addition to the OECD-DAC criteria. These two quality and accountability standards are explored in the following last section of this chapter. With Buchanan-Smith and Cosgrave (2013, pp.52-57) it can be concluded, that the OECD-DAC criteria are not perfect and might remove the focus on the evaluation questions, but there are good reasons for using them instead of devising new criteria for each evaluation:

- Using the OECD-DAC criteria makes drawing lessons from a wide range of evaluations (meta-evaluation) easier
- The OECD-DAC criteria are likely to capture common weaknesses in humanitarian action, based on experience and research
- Using standard criteria makes evaluations easier for both, evaluation managers and evaluators

2.4 Quality and Accountability Standards in Humanitarian Response and WASH

Quality and accountability standards are part of an external regulatory environment for humanitarian action. Their origins can be traced back to the *Code of Conduct for the International Red Cross and Red Crescent Movement and Non-Governmental Organisations*

(NGOs) in Disaster Relief (see e.g. The Sphere Project, 2011, pp.368-376), adopted in 1994. Since its inception, the code has developed a momentum for further so called *Quality and Accountability Initiatives* (Raynard, 2000, p.13). Most of them primarily aim at setting standards for humanitarian work. It is important to recognise that humanitarian standards have come a long way and each of the set of standards has been conceived against the background of specific weaknesses in humanitarian response. They largely (considering the most influential ones) build on previous initiatives and complement each other. Out of the many standard initiatives founded in the past two decades, the leading initiatives from a global perspective are:

- **People In Aid** (established in 1995)

  Focus: Enhancing the quality of human resources and people management
- **The Sphere Project** (established in 1997)

  Focus: Ensuring minimum standards in key areas of humanitarian response
- **Humanitarian Accountability Partnership** (established in 2003)

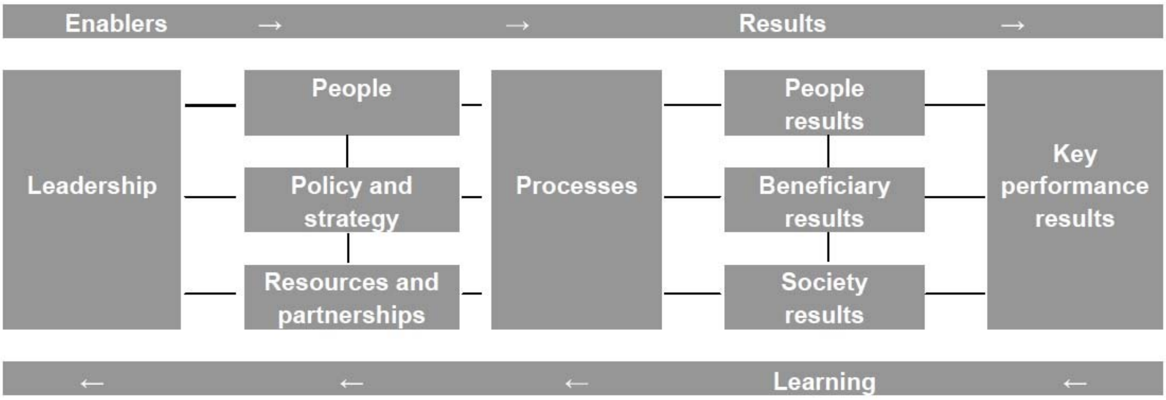
  Focus: Promoting accountability in humanitarian action

These three initiatives are currently pioneering together in the *Joint Standards Initiative (JSI)* to develop a coherent standard architecture (the first tangible outcomes can be expected in 2014). The underlying reason for this collaborative effort is rather pragmatic. As stated on the JSI's website (*http://www.jointstandards.org/about*): *"Whilst the early 1990's saw an absence of standards, the current situation may pose the opposite problem, with at least seventy standards initiatives now in existence in the humanitarian sector."*. For the most part, the quality and accountability mechanisms developed so far are elaborated frameworks, and ever since implementing them has not been straightforward for humanitarians. They are often not understood according to their intentions, usually voluntary and relying on self-regulation in the absence of a humanitarian body with a respective mandate for controlling their adherence. Probably their greatest credit can be seen in having established a closer agreement on what constitutes good practice within the humanitarian community.

Because of their scope, focus, and dissemination in the WASH sector, only two of the quality and accountability initiatives are of concern for the research: *The Sphere Project* and the *Humanitarian Accountability Partnership*. Their standards are partially overlapping, but complementary. While Sphere has a much stronger focus on the quality of humanitarian interventions and provides the only WASH specific standards in emergencies, HAP has been called into life to promote accountability to crisis-affected people. *Figure 8* on the next page shows a general model for comparing the quality and accountability elements of existing initiatives. In this model, quality and accountability is defined as follows (The Sphere Project, 2009, p.4):

Quality is about learning what you are doing well and doing it better. It also means finding out what you may need to change to make sure you meet the needs of your service users.

Accountability describes the ways in which organisations and projects involve different groups in making decisions, managing activities, and judging and challenging results.



Source: The Sphere Project, 2009, p.5

Figure 8: Elements of quality and accountability framed in a comparative model

The model illustrated above has been used to compare several quality and accountability initiatives (see The Sphere Project, 2009) and provides a practical way of understanding the main logic behind them. There are four *Enablers* (*Leadership, People, Policy and strategy, Resources and partnerships*) which work together through appropriate *Processes* to drive the *Results* (*People results, Beneficiary results, Society results, Key performance results*) in an organisation or project. The loop closes through *Learning*, i.e. how the organisation or project learns from results in order to improve the way it works (The Sphere Project, 2009, p.6). The *Enablers, Processes and Results* are defined as follows (The Sphere Project, 2009, pp.5-6):

- ***Leadership***

  How agency leaders, including trustees, directors, managers and team leaders, set direction and enable staff to achieve and deliver results; how leaders make sure an organisation is continually learning and improving.

- ***People***

  How staff and volunteers are motivated, supported and rewarded, and their potential developed and released. "People" includes field staff and/or volunteers, including national, international and temporary staff.

- ***Policy and strategy***

  How an organisation turns its mandate, mission and values into strategy based on the needs of beneficiaries and other stakeholders. How it puts strategy into practice through policies, plans, target and processes, including through use of sectoral standards and principles.

- ***Resources and partnerships***

  How an organisation or project works with partners and resources in order to achieve results. Partners can be any external group, organisation or individuals. Resources include money, equipment, technology, goods, information or knowledge.

- ***Processes***

  How an organisation designs and manages organisational and project processes. How it converts "enablers" to ensure that goods and services are delivered effectively, efficiently, in ways that meet the needs and expectations of beneficiaries and other stakeholders.

- ***People results***

  How the organisation or project and its management and support of staff or volunteers are seen by the people working in it.

- ***Beneficiary results***

  How the organisation or project understands and measures what it is achieving for intended beneficiaries, its main stakeholder group.

- ***Society results***

  Results for a project's wider community and other stakeholders such as donors, host community, other government and non-governmental organisations and their view of project impact.

- ***Key performance results***

  How well the organisation or project achieved what it planned to achieve.

In short, this is what is meant with elaborated frameworks. And the HAP and Sphere standards are such frameworks, albeit the latter do not explicitly deal with all of the model's elements (the Sphere standards are less focused on the elements *Leadership, People results and Society results* (The Sphere Project, 2009, p.9)). The comparative model underlines further aspects of complementary between both standards (The Sphere Project, 2009, pp.9-15): At the core of the HAP standard are accountability and quality management processes to enable agencies delivering optimum results; with a slightly different focus, the Sphere standards address core management processes too, but ultimately concentrate on the results for those affected by an emergency, which are achieved through minimum technical standards (such as in WASH), also as means for M&E.

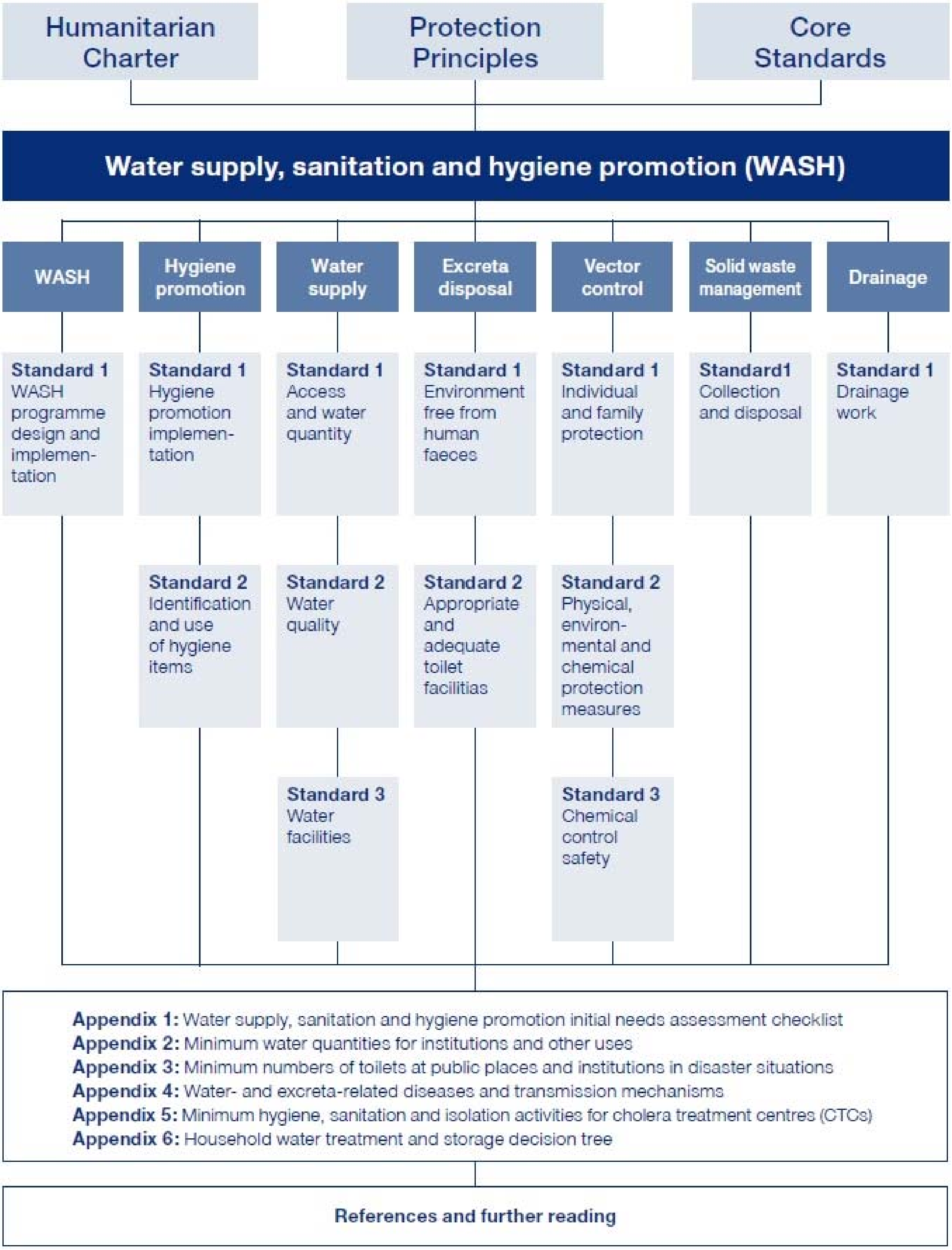
Central to both standards is the emphasis put on the role of crisis-affected people in all stages and at all levels of emergency response. They are designed to be generally and globally applicable in any kind of humanitarian response and have been translated into dozens of languages. But the Sphere and HAP standards are much more than just standards. They are shared expressions of what must be considered at the heart of emergency response based on the wealth of experience gained in the past decades. Furthermore, the standards are continuously evolving and being refined in the light of lessons learned from previous disasters. In particular, the best known quality and accountability resource of *The Sphere Project*, the *Humanitarian Charter and Minimum Standards in Humanitarian Response* (The Sphere Project, 2011) – often referred to as the *Sphere Handbook* – is essentially almost 400 pages of lessons learned and good practice guidance.

The structure of both sets of standards is indispensable to their application. Their common fundament are principles, rights, duties, commitments etc. as expounded in the *Humanitarian Charter* (for the list of key documents that inform the charter see The Sphere Project, 2011, pp.356-367). In the following, first the Sphere standards' structure and then the one of the HAP standard is briefly outlined. From the *Humanitarian Charter*, Sphere derives four basic *Protection Principles,* which together with the charter inform the *Core Standards*. The latter describe processes that are essential to achieving the minimum standards, i.e. in this research, the Sphere minimum standards for WASH. *Figure 9* on the following page depicts them and how they relate to the overall structure.

Because of the standards' structure, and this is the main point that Sphere wants to convey, e.g. the minimum standards in WASH are to be conceived together with the core standards and the protection principles. In relation to the *Humanitarian Charter*, the Sphere Handbook mentions that the WASH minimum standards are "not a full expression of the right to water and sanitation" but reflect the "core content" of it (The Sphere Project, 2011, p.84). Sphere's protection principles are introduced in the last section of *Chapter 3* which deals with cross-cutting themes. The definitions of the core standards as well as the minimum standards in WASH are provided in *Chapter 4*.

Based on several general commitments and principles, the HAP standard is made up of two kinds of requirements that both must be fulfilled to comply with the standard. There are three general requirements and a set of specific requirements which refer to the *HAP Standard Benchmarks* against which performance is assessed. The HAP standard benchmarks (see *Chapter 4* for their definitions) and each of their requirements, i.e. the policies, procedures and practices in place to meet a benchmark (HAP International, 2013, p.26), can be perceived as a checklist an organisation has to go through and act on step by step to meet the standard. Each specific requirement is accompanied with means of verification. These are used for HAP's own certification scheme that includes e.g. on site audits of emergency programmes. In comparison to the Sphere standards, the HAP standard is more concrete about its implementation.

Particularly for promoting accountability, the GWC has been active in publishing materials for field use that build on the Sphere and HAP standards. These include e.g. *WASH Accountability Resources: Ask, Listen, Communicate* (GWC, 2009b), a handbook with down-to-earth good practice examples for implementing accountability; and the appendant *WASH Accountability Checklist* (GWC, 2009a), a draft template which features WASH specific requirements relating to the HAP standard benchmarks. The checklist is attached in *Annex 2*. Finally, humanitarian agencies themselves have developed quality and accountability mechanisms and tools based on one or the other standard initiative. Consequently, the adopted standard is usually part of an organisation's evaluation framework. E.g. the *IFRC Framework for Evaluation* (IFRC, 2011a) includes the Sphere standards as evaluation criteria. On the basis of the GWC and accepted practice it can be argued that the Sphere and HAP standards are fundamental to cover quality and accountability in WASH emergency interventions. Hence, they should be used when evaluating them, regardless of which standard(s) an organisation might have adopted and included in its evaluation policy.



Source: The Sphere Project, 2011, p.82

Figure 9: The structure of the Sphere minimum standards for WASH

3 Using the OECD-DAC Criteria for Evaluations in the WASH Sector

The first part of the meta-analysis of evaluation reports juxtaposes *definition and intended use* for each OECD-DAC criterion with its *applied use* and *key findings* to elaborate *further considerations* regarding its application. The definitions of the OECD-DAC criteria and thoughts on their proper use are based on a quasi-standard ALNAP guide for EHA (Beck, 2006) and supplemented by other EHA guides, manuals and reviews (Cosgrave, Ramalingam and Beck, 2009; Buchanan-Smith and Cosgrave, 2013; DG ECHO, 2007; Buchanan-Smith and Cosgrave, 2010?; Hallam, 1998). The presented findings from the evaluation reports have been extracted with a focus on learning, also considering the WASH gaps (*Chapter 2*), and are therefore referred to as selected key findings in *Table 2-9*. On their basis, further considerations are elaborated with insights from literature and, in particular, meta-evaluation studies of the humanitarian sector's performance in the past years (Stokke, 2007; Harvey, et al., 2010; Taylor, et al., 2013). The last section of this chapter will complete the analysis of the OECD-DAC criteria with answering how important cross-cutting themes are used in conjunction with them.

3.1 Relevance/Appropriateness

3.1.1 Definition and Intended Use

Relevance/Appropriateness is defined as (Beck, 2006, p.22):

"Relevance is concerned with assessing whether the project is in line with local needs and priorities (as well as donor policy)."

"Appropriateness is the tailoring of humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly."

Beck (ibid.) further explains the definition:

"Relevance and appropriateness are complementary criteria that can be used at different levels. Although interventions may be relevant at the macro level, this does not necessarily mean that they are appropriate in terms of the type of activity selected. […] Additionally, the appropriateness of the actual resources or support provided should be evaluated."; the author also suggests that most of the cross-cutting themes (see the last section of this chapter) can be considered under this criterion.

The following issues to consider are highlighted in EHA publications:

- "Analysis of context and an adequate needs assessment are of particular importance for promoting relevant and appropriate responses, and evaluators should pay close attention to the extent to which the planning, design and implementation of interventions takes into account local context." (Beck, 2006, p.23).

- "Evaluators need to pay attention to questions of cultural relativism. For example, in countries with a relatively high standard of living, should interventions be looking to return primary stakeholders to their original condition, or to provide levels of support equal to those in responses in less developed countries? […] There is no easy answer to the question of what constitutes 'need', but it is a question that evaluators should bear in mind when considering the relevance of the response." (ibid.).
- "Evaluations should also evaluate institutional capacity, that is whether there is the capacity in terms of staffing, local knowledge and experience in the country or region, to conduct a relevant and appropriate response" (ibid.).
- "Evaluators should evaluate the extent to which the perceived needs of different stakeholders, in particular women and men, and girls and boys, in the affected population, were met by the intervention." (Beck, 2006, p.24).

3.1.2 Applied Use, Key Findings and Further Considerations

The criteria relevance and appropriateness are used in nine out of ten evaluations. Their intended use – as described above – is well reflected in the evaluation reports. As shown in *Table 2* on the next page, the criteria are used in conjunction with several of the recommended issues to consider. In particular the analysis of the local context and accordingly the quality of needs assessment as well as the resources and capacities to respond to identified needs and regarding the selected approach are dominant subjects. Especially where evaluators discovered an insufficient assessment of the context to inform the response, cross cutting themes (otherwise taken for granted in emergency programming) such as gender and protection are emphasised (Grünewald, et al., 2011a, pp.33-34; Jeene, 2010, pp.11-13). Because assessments play a key role for relevance/appropriateness, identifying and targeting the needs of the vulnerable groups together with participation of affected communities at the outset of emergency interventions are further underlying and interrelated cross-cutting themes in most evaluations (*Table 2*).

Overall, WASH interventions were found to be much more relevant than appropriate. The emergencies whose responses were evaluated have been generally characterised by a mismatch between available resources and needs. Hence, relevance was less of a concern. Given the preventive health character of WASH emergency interventions, even weak needs assessments have less impact on relevance than on appropriateness. The evaluation reports' findings articulate the reasons for less appropriate interventions clearly. The single most noted complex of reasons is lack of community involvement in response design and implementation, which at the same time explains not assessing e.g. pre-existing WASH practices and relates to neglecting affordable community solutions or targeting/meeting needs (*Table 2*). Furthermore, because of these reasons, the appropriateness of a response is causally linked to its connectedness and effectiveness in several key findings. In this regard, additionally a certain lack of flexibility to react on a disaster's changing environment or prolonged nature is stated in evaluation reports. Without doubt, inflexible interventions are partly conditioned by lack of evidence/experience when to best phase out only temporarily appropriate activities and start to invest in solutions that can be maintained locally.

	Relevance/Appropriateness
	Subject/Issue
Criteria used in conjunction with:	Assessment quality and its influence on WASH intervention strategies and overall performance; Gender and protection (Grünewald, et al., 2011a, p.33); Selection of implementing partner organisations based on long-term presence and expertise in both emergency and development; Construction and rehabilitation of WASH infrastructure (Barham, Nabunny and Philpott, 2011a, p.17); WASH sector emergency response model of INGOs in the light of missing government leadership (DeVillez, Bousquet and Nyagwambo, 2011, pp.12-15); Consideration of community-focused solutions and local sanitary and social contexts (Grayel and Mattson, 2012, p.48); Choice of technology and community involvement in response design (Fox, 2008, p.7); Focus/priorities of the intervention and community mobilisation approach; (Fortune and Rasal, 2010, pp.10-13); Context-specific targeting of needs of most vulnerable communities; DRR (Luqman, Zulqarnain and Bahadar, 2012, pp.6-8); Targeting of beneficiaries and technical aspects of WASH approach (Laurens, 2005, pp.40-49); Intervention focus and advocacy; Gender and protection (Jeene, 2010, pp.11-13)
Selected key findings related to criteria:	Poor assessments have overlooked pre-existing local WASH practices and leading to inadequate strategies (Grünewald, et al., 2011a, p.33); Relevance and appropriateness is causally linked with the WASH approach’s effectiveness (rehabilitation of existing infrastructure rather than new construction of facilities); Long-term presence in the region of operation and expertise in both emergency and development is regarded essential for implementing organisations (Barham, Nabunny and Philpott, 2011a, p.17); It is crucial to support re-emerging governmental WASH sector coordination and capacity to safeguard an appropriate response to any new emergency (DeVillez, Bousquet and Nyagwambo, 2011, p.2); While interventions reliant on external assistance (such as water trucking) may be relevant and appropriate in the immediate response phase, offering affordable community solutions appropriate to the local contexts as well as preparing for the transition and withdrawal of activities need to be addressed right from the beginning of a response (Grayel and Mattson, 2012, p.48); Even in emergency response operations with strong governmental leadership and planning, community involvement in the response design is vital to target needs and provide lasting solutions (Fox, 2008, p.7); Only a wider and more flexible response to excreta disposal (beyond a possible appropriate focus on latrine construction) may further reduce unsafe practices; Behavioural details/assumptions (e.g. squat/sitting toilet) based on an incomplete understanding of the context and informing the intervention design impact overall performance strongly (Fortune and Rasal, 2010, pp.10-13); Participatory planning and rigorous assessments need to go hand in hand to ensure relevant targeting and appropriate strategies for restoring livelihoods (integration of WASH with food security and livelihood, and adoption of cash transfer modalities); (Luqman, Zulqarnain and Bahadar, 2012, pp.6-8); More systematic hygiene promotion (e.g. maintaining hygiene with local means) and adaption of technical solutions to cover inaccessible areas and support home-made, low cost durable latrines (e.g. provision of unavailable materials) (Laurens, 2005, pp.40-49); It is important to record needs that cannot be addressed with the own intervention and to lobby other organisations to fill these gaps in the response (Jeene, 2010, pp.11-13)

Table 2: Relevance/Appropriateness: Summary of use and key findings

The key findings in *Table 2* further underline that strong governmental leadership and response planning or otherwise working through local organisations and their structures do not automatically lead to meaningful involvement of disaster-affected communities or appropriate intervention designs. Overall, the findings from the evaluation reports are unsurprising in the light of on-going meta-evaluations of the humanitarian sector and do not add any new evidence to known problems and weaknesses. E.g. Stokke (2007, pp.13-14) points out the need for mapping and monitoring needs and target groups to ensure appropriate strategies and allocation of resources; similarly, the two latest *State of the Humanitarian System* reports (Harvey, et al., 2010, pp. 27-30; Taylor, et al., 2013, pp.49-52) keep finding that primary stakeholders are not adequately consulted about their needs and how these could best be met as a crucial first step to assure targeted emergency projects with best possible quality.

Precisely because such findings are not new at all, the Sphere and HAP standard initiatives have been launched. These standards encompass the core aspects of relevance and appropriateness in their definitions and moreover provide a set of accompanying measures that would allow evaluators to explore the underlying reasons for lack of community involvement and related general findings (measures include the key actions, key indicators and guidance notes of the Sphere standards (The Sphere Project, 2011, pp. 49-138) and the requirements to meet the HAP standard benchmarks (HAP International, 2013, pp.24-51)). Of course the evaluation reports offer much more details on e.g. why interventions have not been appropriate than can be presented in this analysis but it appears cumbersome and less productive to reveal through relevance/appropriateness mainly what has been the starting point for humanitarian standards.

Above considerations question if these criteria are still relevant and appropriate for their use in combination with other OECD-DAC criteria, cross-cutting themes and standards, for both accountability- and learning-oriented evaluations. The Sphere and HAP standards constitute alternative evaluation criteria which should be included anyway. Although several evaluations include selected issues these standards are concerned with under the heading relevance/appropriateness, their notion is sparse and incoherent. Finally, for creating evidence that a given form of response is the most appropriate one in a particular context, standards are indispensable. To assess appropriateness, evaluations almost exclusively depend on the views of crisis-affected people, but as mentioned earlier, they are usually not more than informants. The measures and indicators provided with the Sphere and HAP standards could as well be utilised to increase their level of participation in evaluations. In fact, one of the intentions of the HAP standard is to promote participatory evaluations (HAP International, 2013, pp.38-41) and the Sphere handbook too is designed for M&E by non-experts in this regard.

3.2 Connectedness

3.2.1 Definition and Intended Use

Connectedness is defined as (Beck, 2006, p.27):

“Connectedness refers to the need to ensure that activities of a short-term emergency nature are carried out in a context that takes longer-term and interconnected problems into account.”

Beck (2006, p.28) further explains the definition:

“Evaluators should concentrate in particular on whether the key linkages between the relief and recovery phases have been established, for example the existence of a sound exit strategy with timelines, allocation of responsibility and details on handover to government departments and/or development agencies, and adequate availability of funding post-response. Evaluation of connectedness is often linked to evaluation of impact, as both tend to consider longer-term consequences.”

The following issues to consider are highlighted in EHA publications:

- “Evaluators should be sensitive to relative expenditure on relief and recovery. […] Recovery activities should be evaluated against appropriate indicators, which may not be the same indicators used to evaluate relief” (Beck, 2006, p.28).
- “Connectedness and partnerships. […] Evaluators should analyse the nature of partnerships supporting connectedness, for example how they came into being and were supported, so that others can learn from this experience.” (ibid.).
- “Evaluations should examine the extent to which local capacity is supported and developed. Development of and support to capacity is central to ensuring that the effects of interventions are not lost. For example, a number of evaluations have pointed out that establishing capacity of water-user associations to manage water facilities may be too great a challenge for the relief phase […]. Evaluations should also examine the degree to which livelihoods of the affected population are supported or disrupted by the intervention, as this will have a significant impact on longer-term results. They should also analyse the degree to which the capacity of government, civil society and other partners at various levels is built by the intervention.” (Beck, 2006, pp.28-29).
- Connectedness should also address the cross-cutting concerns of environment and DRR (Cosgrave, Ramalingam and Beck, 2009, p.18).

As the criterion *Connectedness* has been deduced from *Sustainability* – “the idea that interventions should support longer-term goals, and eventually be managed without donor input” (Beck, 2006, p.27) – evaluation findings referring to *Sustainability* in place of *Connectedness* will be conceived with the latter.

3.2.2 Applied Use, Key Findings and Further Considerations

The criterion connectedness respectively sustainability is used in nine out of ten evaluations (four evaluations employed sustainability instead of connectedness). Although connectedness is better suited for EHA than sustainability as there is “no consensus concerning the *extent* to which humanitarian action should support longer-term needs” (Beck, 2006, p.27), which criterion is used makes no difference in the researched evaluations in terms of the intended use of connectedness. Both criteria or rather the reasons for their use are well understood and even heavily emphasised in some evaluation reports which found interventions to have neglected longer-term and interconnected problems (e.g. Fortune and Rasal, 2010, p.29). As subsequent *Table 3* shows, the key issue connectedness is used in conjunction with, is operation and maintenance of WASH facilities. More or less all evaluations refer to it which includes a whole range of aspects: the effect of upcoming emergencies, governmental WASH policies, cost recovery/financing, community ownership, trainings, spare part supply or land property conflicts. Other concrete subjects addressed in the evaluation reports are for instance the need for exit indicators in complex contexts, integrating LRRD in general operations management, capacity building and the importance of pre-existing partnerships/relations between INGOs and local organisations/institutions.

Because the positive health impacts of WASH interventions can only be realised if services continue over time, connectedness is a key criterion that should be used in any evaluation. According to the key findings (*Table 3*), the reasons or risks for discontinuation of WASH services after the end of emergency projects primarily relate to the difficulties of establishing community based operation and maintenance schemes for WASH facilities. As *Table 3* illustrates, some of them are easier to address through involvement of targeted communities already in the response design (e.g. considering household connections as incentives to pay for water), while others are complex and require humanitarian actors to think of more than just WASH (e.g. livelihoods/income generation or supporting commercial vendors in establishing a supply chain for spare parts). The latter goes clearly beyond a focus on building back better and entails humanitarian response to be based on transparent and realistic exit strategies that allow cluster/sector coordination mechanisms and their actors (state authorities, other organisations) to build on.

The key findings highlight the importance of considering exit strategies from the beginning of a response, i.e. when decisions for certain intervention approaches and technologies are made. In this respect, connectedness becomes central for RTEs. The evaluation of Fortune and Rasal (2010) during the Haiti earthquake response provides a good example of how early decisions (on e.g. elevated pit latrines, raised latrines with septic tanks and/or urine diversion or household level excreta disposal in plastic bags) and their appropriateness in the light of the context and needs affect the development and realisation of exit strategies. A major weakness addressed in many evaluation reports is still unclear, fragmentary or too late implemented exit strategies with little relevance. The establishment of exit indicators in murky contexts (Barham, Nabunny and Philpott, 2011b, p.14) is a further aspect of it, and which points towards the difficulty of evaluating when agencies can or should withdraw from areas where the Sphere minimum standards in WASH are unattainable.

	Connectedness
	Subject/Issue
Criterion used in conjunction with:	LRRD in view of new emergencies (Grünewald, et al., 2011a, p.34); Financing operation and maintenance of WASH facilities; Exit indicators for complex/murky contexts (Barham, Nabunny and Philpott, 2011a, pp.20-21; Barham, Nabunny and Philpott, 2011b, pp.12-17); Allocation of means and government policies (Grayel and Mattson, 2012, p.48); Cost recovery, management, ownership and human resources aspects of operation and maintenance of water plants (Fox, 2008, pp.15-22); Patterns of conflict in society and operation management (Fortune and Rasal, 2010, p.29); Good practice demonstration/spill over effects and engagement of affected communities with local authorities (Luqman, Zulqarnain and Bahadar, 2012, p.28); Ownership of and community responsibility for water points; Training, maintenance and spare part supply (Laurens, 2005, pp.49-62); WASH facilities maintenance and land property conflicts (Van der Wijk, Mwezi and Kazan, 2010, pp.23-24); Pre-existing partnerships/relations between INGOs and local organisations/institutions; Integration with community and government structures; Financing WASH infrastructure maintenance (Jeene, 2010, pp.13-14)
Selected key findings related to criterion:	The short duration of projects and new emergencies occurring during the response limit the impact of community approaches and require rethinking exit strategies (Grünewald, et al., 2011a, p.34); Community based operation and maintenance schemes for WASH facilities will only work in stable contexts (regarding livelihoods and public health); There is a need to establish exit indicators for withdrawing activities to provide a degree of clarity in complex/murky contexts (Barham, Nabunny and Philpott, 2011a, pp.20-21; Barham, Nabunny and Philpott, 2011b, pp.12-17); It is crucial to design, implement and monitor WASH exit strategies from the early stages of the response, based on needs, opportunities and constraints (Grayel and Mattson, 2012, p.55); Providing incentives for users to pay for water based on their uses and ideas (e.g. household connections) and creating ownership of infrastructure and operational management of services would increase the sustainability (Fox, 2008, pp.7-20); Considering pre-existing social conditions for the intervention is indispensable to ensure proper long-term planning; Frequent turnover of staff is likely to represent an unnecessary discontinuity in achieving outcomes (Fortune and Rasal, 2010, pp.28-29); Construction/rehabilitation of WASH related infrastructure will only address longer term vulnerabilities if maintenance by communities is addressed; More efforts are required to establish a thorough understanding of disaster/conflict-affected communities in various risk reduction measures at the local level (Luqman, Zulqarnain and Bahadar, 2012, p.39); High level of community ownership and participation are important but need to be complemented by technical aspects of maintenance (e.g. transparency of fee collection, accounting and hand pump maintenance trainings, creation of spare part stocks and shops) (Laurens, 2005, pp.49-62); Livelihoods and resilience of affected people need to be addressed to make cost recovery schemes work; Land owners on whose property WASH infrastructure is erected and who do not benefit from projects present a risk (Van der Wijk, Mwezi and Kazan, 2010, pp.23-24); Insufficient government capacities hampering the integration of interventions and their structures should be taken into account in the response design; Long-term spare part supply is best done by the private sector, i.e. commercial vendors (Jeene, 2010, pp.13-18)

Table 3: Connectedness: Summary of use and key findings

For this complex of problems e.g. the *ACF International Network* has published the practical manual *How to make WASH projects sustainable and successfully disengage in vulnerable contexts* (ACF IN, 2008), which may also provide valuable pointers for evaluators when considering connectedness. It builds on the lessons learned from past WASH interventions including for instance the evaluation of Laurens (2005) and identifies the main reason why projects become unsustainable not to be technical issues but related to "management, social relationships and community dynamics" (ACF IN, 2008, p.6). This should also be seen against the backdrop of "growing physical and social detachment of many international aid personnel from the societies in which they work" because of increasing security concerns and a "substantial shift towards 'remote management' techniques […] through national and local field workers, subcontracted intermediaries and new technologies" (Collinson and Duffield, 2013, p.iii). The point is that decision-making power is located within remote management and the vast majority of national/local project staff should not hide the fact that hierarchies of inequalities define the humanitarian system (Collinson and Duffield, 2013, p.2) and lie at the heart of weakened control of relief delivery chains and distorted emergency programming (Collinson and Duffield, 2013, pp.24-25).

Evaluators need to be aware of these underpinnings to not draw rash conclusion such as mistaking empowering disaster-affected communities as the key determinant for connectedness without considering other factors. Meta-evaluations highlights capacity building and long-term support of national actors as a continued weakness of the humanitarian system, in particular with respect to ensuring that coordination of interventions is "tailored to support state response capacities" (Taylor, et al., 2013, pp.68-73). Indeed, WASH interventions that would primarily focus on supporting existing NGO and state capacities (where this is not contradicting humanitarian principles) would look quite different than the ones described in the evaluation reports. All of the above, although not being exhaustive, can be combined under LRRD. Some aspects of it are part of the Sphere and HAP standards (e.g. exit strategies or support of existing capacity), however, as for its complexity the criterion connectedness is best suited to explore the wide range of potential issues involved. Finally, as *Table 3* occasionally demonstrates, connectedness also serves as an important bracket with which findings related to other OECD-DAC criteria can be viewed in conjunction with standards and cross-cutting issues such as DRR and environment.

3.3 Coherence

3.3.1 Definition and Intended Use

Coherence is defined as (Beck, 2006, p.33):

"The need to assess security, developmental, trade and military policies, as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human-rights considerations."

Beck (2006, pp.33-34) further explains the definition:

"Assessment of coherence should focus on the extent to which policies of different actors were complementary or contradictory. […] Coherence becomes an important evaluation issue when politics fosters the occurrence or continuation of a humanitarian emergency, and when military and civilian actors are involved in the same emergency […] Coherence can also be analysed solely within the humanitarian sphere, to assess whether all actors – including governments, the UN, donors, civil society and the private sector – are working towards the same goals. As the definition makes clear, evaluation of coherence needs to take into account considerations of humanitarian space, including protection. […] Evaluations need to consider whether actors have been coherent in their approach to protection, and whether policy has met the protection needs of primary stakeholders."

The author (ibid.) states that coherence has proved the most difficult of the OECD-DAC criteria to operationalise and is often confused with coordination; he further argues, the "evaluation of coherence should complement that of coordination, with coherence focusing on the policy level, and evaluation of coordination focusing more on operational issues". Coherence may also be less relevant to consider in "single-agency, single-project evaluations, but if an evaluation is not going to consider policies as benchmarks against which to measure results, the reasons for this should be made clear" (ibid).

The following issues to consider are highlighted in EHA publications:

- Coherence is one of the most political OECD-DAC criteria to evaluate and requires specific capacity and resources of the evaluation team (Beck, 2006, p.34).
- With increased involvement of military forces, "evaluators need to pay close attention to the mandates, agendas and principles of different actors", and evaluate also if there should be coherence at all (Beck, 2006, p.35).

3.3.2 Applied Use, Key Findings and Further Considerations

The criterion coherence is used in six out of ten evaluations. As following *Table 4* shows, the confusion with coordination is evident in two evaluations (Luqman, Zulqarnain and Bahadar, 2012; Jeene, 2010). In the other four evaluations, coherence is used to assess donor funding against its own policies and agencies' compliance with LRRD strategies, WASH Cluster/government standards and government development plans. Because the evaluations commissioned by implementing organisations were classic single project/programme evaluations focussing on their own work, the full potential of this criterion could not be exploited. By definition, coherence has a strong accountability focus and is ideally used in situations where multiple actors are expected to base their interventions on a shared strategy. Although WASH clusters form a dominant part of the response in most of the analysed emergency contexts, such strategies are less common or difficult to achieve in a progressive and comprehensive manner. Only one emergency WASH project was designed on the basis of a clearly elaborated strategy, however the evaluation (Fox, 2008) did not explicitly employ the criterion coherence and subsumed its aspects under sustainability.

The key findings of those evaluations in which coherence was used according to its intentions (*Table 4*) demonstrate the importance of sector collaboration for working out and implementing transition respectively LRRD strategies as shared strategies which stakeholders are committed to. Also, organisations' transparency about challenges and achieved results and sustainability (in the meaning of longevity) of technology are deemed as important in this regard. Fox (2008, p.19) emphasises that government response plans should be reviewed by implementing organisations according to their considerations given to core humanitarian standards. Meta-evaluation reports (Harvey, et al., 2010, pp. 44-46; Taylor, et al., 2013, pp.79-81) note that for humanitarian organisations maintaining independence is becoming increasingly difficult where aid is politicised and that in some country contexts of the evaluations such as Pakistan or DRC, political and military considerations as well as government and UN policies are in direct conflict with humanitarian goals. Taylor, et al. (2013, p.81) additionally mention a "growing divide between 'strongly principled' humanitarian agencies and the others".

Coherence	
	Subject/Issue
Criterion used in conjunction with:	Donor funding of present and future interventions (Barham, Nabunny and Philpott, 2011b, p.2); How the response approach fits with LRRD strategies (DeVillez, Bousquet and Nyagwambo, 2011, pp.21-24); Following guidance and standards of government agencies and the WASH Cluster (Grayel and Mattson, 2012, p.49); If intervention is in-line with government development plans and goals (Fox, 2008, p.19); Avoidance of duplications in response and cluster coordination (Luqman, Zulqarnain and Bahadar, 2012, p.39); Links of the intervention to other sectors and development (Jeene, 2010, p.14)
Selected key findings related to criterion:	A shared analysis of the development challenges is essential to the emergence of a coherent transition strategy (Barham, Nabunny and Philpott, 2011b, pp.7-17); "A key component of any process which attempts to link relief to rehabilitation and development is the sustainability of the technology which was installed during the emergency response phase of the relief exercise."; A coordinated transition strategy involving all stakeholders is essential (DeVillez, Bousquet and Nyagwambo, 2011, pp.21-24); Humanitarian organisations need to be "more transparent about actual coverage achieved as well as changes in coverage over time due to population shifts and increase or decrease in access to services" (Grayel and Mattson, 2012, p.49); Even in case of strong government leadership in emergency response it is important to reflect if the intervention approach based on government plans is adequate and responds to defined needs and capacities of the affected communities (Fox, 2008, p.19); Underlying/existing vulnerabilities make an emphasis on community based DRR approaches essential (Luqman, Zulqarnain and Bahadar, 2012, p.39); A coherent intervention needs dedicated staff and clear responsibilities for such issues (Jeene, 2010, p.14)

Table 4: Coherence: Summary of use and key findings

Above findings suggest that coherence or rather the reasons for non-coherence can be particularly useful to assess in politicised aid context, especially in conjunction with connectedness. However, the researched evaluation reports do not indicate any such context or that intervention goals have been in opposition to government, military or UN policies. The purposes for which coherence is used in them is better captured with connectedness respectively the Sphere and HAP standards. As the evaluation of Fox (2008) illustrates, it appears unaptly to not assess issues related to LRRD under connectedness and unnecessary to not review response planning of governments and policies in general directly on the basis of humanitarian standards. The evaluation reports and other universal WASH publications such as the very comprehensive *Model Guidelines for Mainstreaming Water and Sanitation in Emergencies, Protracted Crises, LRRD and Disaster Preparedness Operations*[3] (DG ECHO, 2005a) do not indicate any other realms than those of connectedness or humanitarian standards for which coherence is primarily concerned with in WASH. Of course, politicised emergency response operations can be an exemption in which coherence may be used to reveal valuable lessons learned. After all, sector-wide joint evaluations, though rare, would be an ideal setting for considering the criterion coherence.

3.4 Coverage

3.4.1 Definition and Intended Use

Coverage is defined as (Beck, 2006, p.38):

"The need to reach major population groups facing life-threatening risk wherever they are."

Beck (ibid.) further explains the definition:

"Evaluation of coverage involves determining who was supported by humanitarian action, and why. In determining why certain groups were covered or not, a central question is: 'What were the main reasons that the intervention provided or failed to provide major population groups with assistance and protection, proportionate to their need?' […] Coverage is linked closely to effectiveness, as objective statements in humanitarian action, which are assessed with reference to effectiveness, often refer to numbers or percentages of the population to be covered."

The following issues to consider are highlighted in EHA publications:

- Evaluators should consider "whether aid was provided according to the need", internationally, nationally and locally (Beck, 2006, p.40).
- "At the regional and local levels, evaluators need to assess the extent of inclusion bias, that is inclusion of those in the groups receiving support who should not have been (disaggregated by sex, socioeconomic grouping and ethnicity); as well as the extent of exclusion bias, that is exclusion of groups who should have been covered but were not (disaggregated by sex, socioeconomic grouping and ethnicity)." (Beck, 2006, p. 39).

[3] These model guidelines have been commissioned by ECHO to fill the gap in specialised literature on WASH in humanitarian settings and are based on then (2005) existing WASH sector publications and evaluations as well as extensive interviews and field research at headquarters and operational level. They present "lessons and good practice relating to a range of areas of intervention, including technical, institutional, and social aspects of service provision, as well as hygiene promotion and behavioural change" (DG ECHO, 2005a, p.1). See also the corresponding concept paper (DG ECHO, 2005b) for the main rationale of the model guidelines.

- "Coverage should also include the cross-cutting themes of gender, social exclusion and access." (Cosgrave, Ramalingam and Beck, 2009, p.19).
- "Coverage alerts evaluators that [complex] emergencies and associated humanitarian programmes can have significantly differing impacts on different population sub-groups, whether these are defined in terms of ethnicity, gender, socio-economic status, occupation, location (urban/rural or inside/outside of a country affected by conflict) or family circumstance (e.g. single mother, orphan)." (Hallam, 1998, p.54).
- Evaluators need to be sensitive to differing perspectives on what constitutes need and determine whether targeting practice is appropriate from the perspective of affected people (Beck, 2006, p.40).

3.4.2 Applied Use, Key Findings and Further Considerations

The criterion coverage is used in six out of ten evaluations. It is quite self-explanatory and used in reference to the Sphere minimum standards in WASH (*Table 5*). In fact, its intended use as well as the other issues coverage is linked with in the evaluation reports (i.e. covering major population groups, most urgent needs and the most vulnerable) are well reflected in the Sphere standards (see *Chapter 4* or rather the full conception of the standard by The Sphere Project (2011)). The key findings in *Table 5* mention lack of data (collection) as a severe obstacle for evaluating coverage. Although data disaggregation is not identified as a problem by evaluators, yet in only few evaluation reports disaggregated data is used to back findings. In the evaluation reports, the problems faced in increasing or reaching planned coverage are linked with the appropriateness and effectiveness of emergency response; e.g. pre-disaster context analysis, post-disaster settlement patterns and the built upon intervention design are noted as interrelated issues in *Table 5*.

Meta-evaluation reports highlight that humanitarian needs or rather the stated requirements through appeals outstrip funding by roughly one third between 2007 and 2010 (Harvey, et al., 2010, p. 23; Taylor, et al., 2013, p.43). In this period, overall funding coverage of humanitarian WASH interventions remained just under 50 percent but with significant differences when comparing emergency response operations to different crises (Taylor, et al., 2013, p.45). In the WASH sector, lack of funding is seen as a "major constraint to effective response" and most constrained with respect to "preparedness and risk reduction" (Harvey, et al., 2010, p. 25). This makes the search for more efficient and effective interventions also under the heading *value for money* a top priority. Furthermore, Harvey, et al. (2010, p. 24) find that "[e]valuations tend to focus on the strengths and weaknesses of the aid provided by aid agencies and neglect the question of population that were not reached". Hence, the authors argue, existing evidence does not allow judgements about how adequately needs are met (ibid.). The evaluation reports used for this research are no exception on that score.

Usually, humanitarian actors are quick with announcing quantitative figures on reached coverage as a result of particular interventions. But what matters most is the quality of coverage, i.e. under which conditions services are accessible for a specific number of affected people. The Sphere minimum standards in WASH provide the necessary tool to determine coverage in both quantity and quality. Their proper application would also reveal much more information about the extent of unmet needs. Moreover, the key actions, key indicators and guidance notes accompanying each standard definition contain a level of detail which is better suited for evaluating coverage, which ultimately always involves quantitative measurements, than the definition and intended use described above. Finally, the definition of coverage is also better viewed under effectiveness (one may add efficiency too) as coverage on its own is less meaningful – a practice already apparent in the majority of evaluation reports.

Coverage	
	Subject/Issue
Criterion used in conjunction with:	If the intervention is targeting the most vulnerable people and achieving humanitarian standards (e.g. Sphere minimum standards in WASH) (Grayel and Mattson, 2012, pp.48-49); If planned coverage was reached and adhered to the Sphere minimum standards in WASH (Fox, 2008, p.14); Adherence to the Sphere minimum standards in WASH (Fortune and Rasal, 2010, pp.4-22); If intervention is covering most crucial needs and most vulnerable people (Luqman, Zulqarnain and Bahadar, 2012, p.39); Whether or not the major population groups, including the most vulnerable, have been reached and provided with assistance proportionate to their needs; Sphere minimum standards in WASH (Van der Wijk, Mwezi and Kazan, 2010, pp.22-23); Geographical areas; Biases in the intervention; Covering for other INGOs insecure and therefore inaccessible areas; Sphere minimum standards in WASH (Jeene, 2010, pp.14-16)
Selected key findings related to criterion:	Lack of data collection throughout the intervention makes assessing targeting practices and their results impossible, beyond a rather meaningless number of people reached (Grayel and Mattson, 2012, pp.48-49); (Planned) Coverage and thereby the usage and value of a system (e.g. water purification units) can only be assessed if it is producing according to needs (Fox, 2008, p.14); Coverage is closely linked to lack of pre-disaster context analysis and resulting inappropriate intervention design (Fortune and Rasal, 2010, p.12); The diversity of needs and the resources of the affected people are key determining factors for increasing coverage (Luqman, Zulqarnain and Bahadar, 2012, p.39); Using the Sphere minimum standards in WASH for planning interventions has to be complemented by assessing the actual use of constructed infrastructure (Van der Wijk, Mwezi and Kazan, 2010, pp.22-23); Different post-disaster settlement patterns impact and link both coverage and effectiveness of WASH interventions (Jeene, 2010, pp.14-16)

Table 5: Coverage: Summary of use and key findings

3.5 Efficiency

3.5.1 Definition and Intended Use

Efficiency is defined as (Beck, 2006, p.44):

"Efficiency measures the outputs – qualitative and quantitative – achieved as a result of inputs. This generally requires comparing alternative approaches to achieving an output, to

see whether the most efficient approach has been used."

Beck (ibid.) further explains the definition:

"Efficiency measures how economically inputs (usually financial, human, technical and material resources) were converted to outputs. Assessment of efficiency tends to start with financial data, and should factor in the urgency of assessed needs of the affected population. Determining whether an intervention was implemented in the most efficient way may involve comparison with alternatives […] Efficiency links to other criteria including appropriateness of choice of intervention, and effectiveness"

The following issues to consider are highlighted in EHA publications:

- "Political priorities of governments and agencies may cause interventions to be inefficient." (Beck, 2006, p.45).
- "Part of the assessment of efficiency considers whether goods/inputs were purchased most efficiently in relation to source of input. One key question is whether inputs were locally purchased or imported." (ibid.).
- "Financial areas to consider are: total cost of the intervention broken down by sector; costs of inputs locally and internationally; transportation costs broken down by sector and type of transportation; staff costs, broken down by local and expatriate staff; and administration costs as a percentage of intervention costs." (ibid.).
- Focus on the wider sense of efficiency, i.e. the "quality of process" (DG ECHO, 2007, p.53).

3.5.2 Applied Use, Key Findings and Further Considerations

The criterion efficiency is used in eight out of ten evaluations. *Table 6* (on the next page) shows the range of issues that affect the efficiency of a response according to the evaluation reports. Particularly the intervention design and general operations management, which includes the allocations of resources (financial, personnel, material and logistics) and implementation processes such as coordination/cooperation with other stakeholders, are the most frequently noted set of issues the criterion is used in conjunction with. These underline that efficiency is linked with appropriateness (intervention design) and effectiveness (management of resources and processes). Further, it is worth noting that at least two evaluations emphasised the quality of the conducted work explicitly (DeVillez, Bousquet and Nyagwambo, 2011; Jeene, 2010). Despite some scattered evidence of considering alternative options as more efficient, all evaluations generally fall short in actually comparing the interventions carried out with alternative approaches. Last but not least, financial areas remain largely unaddressed, and where addressed, the value for money analysis is rather superficial and offers little insights through figures. A notable exception in this regard is the evaluation of Laurens (2005).

	Efficiency
	Subject/Issue
Criterion used in conjunction with:	Value for money; Appropriate technology; Scope of intervention design (Barham, Nabunny and Philpott, 2011a, pp.17-18); Implementation and coordination processes; Use of available resources; Quality of interventions (DeVillez, Bousquet and Nyagwambo, 2011, pp.17-21); Intervention design; Collaboration/coordination; Human resources; Knowledge management; Support of local/national capacities (Grayel and Mattson, 2012, pp.49-50); Management of human resources; Operational resources; Information management (Fortune and Rasal, 2010, pp.7-26); Intervention design (Luqman, Zulqarnain and Bahadar, 2012, p.40); Relative intervention expenditures; Logistical arrangements; Human resources; Security; Cooperation with other agencies and local authorities; Monitoring (Laurens, 2005, pp.63-68); Suppliers of WASH materials; Material and personnel costs (Van der Wijk, Mwezi and Kazan, 2010, pp.20-21); Human resources; Quality of work (Jeene, 2010, pp.17-18)
Selected key findings related to criterion:	Technology alien to community life will not survive; Size and nature of systems needed for large scale camps (e.g. motorised water schemes) will likely be inappropriate after the closure of camps; Not considering environmentally-friendly technologies such as rainwater harvesting is a lost opportunity; Good workmanship is crucial (Barham, Nabunny and Philpott, 2011a, pp.17-18); Lack of data (monitoring and reporting) makes an objective assessment of efficiency infeasible; While a “quick and dirty” strategy may be tolerable in the very first phase of an emergency, in later stages, where time and resources are available, poor quality work will only result in inefficiencies having a detrimental effect on the sustainability of installed facilities (DeVillez, Bousquet and Nyagwambo, 2011, pp.17-21); While expensive activities (e.g. water trucking, latrine desludging) can be justified for the very first phase of the emergency, their continuation will be extremely budget consuming, thereby minimising resources for more appropriate or other interventions, and affecting overall performance negatively (Grayel and Mattson, 2012, pp.49-50); Proper documentation and information management, well managed operational means (e.g. warehouse, transportation), utilising local resources (e.g. mobilising the affected community to dig latrine pits instead of importing diggers with limited range of use) and productive division of labour are underlying key factors for efficiency (Fortune and Rasal, 2010, pp.7-26); Cash for work schemes together with unconditional cash grants are efficient approaches to target the most vulnerable people and to rehabilitate and construct household/community WASH infrastructure (Luqman, Zulqarnain and Bahadar, 2012, pp.19-22); Keeping a decentralised WASH approach (e.g. local staff, local knowledge) can reduce logistical constraints and allow local support to be more easily obtained if the security situation deteriorates; Cooperation with other agencies can reveal extremely useful if mutual areas of interventions and methodologies are shared (e.g. sharing experience with spare part supply, training of mechanics, latrine construction, etc.) (Laurens, 2005, pp.63-68); Financial efficiency can be increased when working with local organisations and through community structures and involving beneficiaries in construction and transport of materials (Van der Wijk, Mwezi and Kazan, 2010, pp.20-21); When contracting companies to e.g. rehabilitate boreholes it is important to not only look at the costs but also at their local track record to determine the quality of work previously done and to supervise; Poor quality workmanship (e.g. repairing rather than rehabilitating boreholes) may render the whole intervention ineffective (Jeene, 2010, pp.17-18)

Table 6: Efficiency: Summary of use and key findings

The key findings of *Table 6* echo a range of common pointers on how to render WASH interventions more efficient, thereby increasing effectiveness: Using appropriate, reliable and known technologies, setting value on good workmanship, utilising local resources, early withdraw of comparably expensive activities and using such only to gain time for implementing longer-term solutions (if the latter is what should be achieved), considering cash for work and unconditional cash grants schemes, or finally, ensuring coordination and inter-sector/inter-agency learning to avoid duplications of efforts and mistakes are noted. On the one hand, some evaluation reports (e.g. Grayel and Mattson, 2012) really go into details and provide a sound elaboration of findings which in turn can be considered an acceptable evidence base for learning. On the other hand, however, the main problem in evaluating efficiency is still a lack of (monitoring) data. DeVillez, Bousquet and Nyagwambo (2011, p.18) admit that their assessment of efficiency "during the early part of the period under scrutiny is necessarily rather subjective".

The meta-evaluation of Harvey, et al. (2010, p.43) finds that on the whole, the question of efficiency and cost-effectiveness is "clearly an area where more focus is needed" and identifies corruption risks as a neglected element of efficiency in humanitarian work that is "rarely explicitly analysed, monitored and evaluated in the public domain". Out of the researched evaluations only one (Laurens, 2005, pp.54-56) draws attention towards the risk of corruption in terms of lacking financial transparency and its high potential to cause severe negative impacts on the entire project. The research of Taylor, et al. (2013, p.74-76) particularly emphasises the open questions about value for money in humanitarian action, which is essentially inscribed in the definition of efficiency but certainly includes effectiveness too. The authors (ibid.) note that with increasing interest from donors in efficiency, "significant scepticism, confusion and resistance" of humanitarians about/to the concept *value for money* remains. It is argued that a "focus on efficiencies for comparative purposes simply doesn't work in a context like South Sudan" where any activity is relatively expensive and takes much more time compared to other countries; and others feel that the concept itself "runs counter to the humanitarian ethos" (Taylor, et al., 2013, p.75-76).

In general, it can be concluded that despite anything in humanitarian action being context-sensitive, value for money is still important. It simply matters, when for instance extremely budget consuming activities aiming at rapid WASH service delivery (e.g. water trucking) are carried on for too long, therefore leaving much less money for interventions that benefit the targeted population in the longer term. And it should be acknowledged that several of the more or also quality related aspects of efficiency (at least human, technical and material resources) are very well suited for comparative purposes. For example, Grünewald, et al. (2011b, p.42) state for disasters in urban settings that the "typical profile of a WaSH NGO worker is often not appropriate and it is necessary to explore other rosters offering electrical engineers and water network specialists". According to Stokke (2007, pp.14-16), much more should be learned about the synergies between local, national and international capacities – a cornerstone of efficient emergency response.

To sum up, efficiency is a core evaluation criterion under which establishing credible evidence is likely to be hampered in varying degrees by lack of data, but whose purpose serves more than a sometimes purely economic perceived value for money analysis –

particularly for learning. Probably the best example to illustrate the importance of efficiency for evaluating innovative approaches is the rise of cash-based programming in emergencies. Although having a long history dating back to the 19th century, only in the last decade its perception has shifted from “radical and risky” to a “mainstream programming approach” (Ramalingam, Scriven and Foley, 2009, pp.43-44). Case studies and evaluative research have prepared the ground for this “paradigm-shift in how humanitarian assistance is undertaken (in relation to the perception of beneficiaries as active participants)” (ibid.). As the evaluation of Luqman, Zulqarnain and Bahadar (2012, pp.19-22) demonstrates, cash-based WASH interventions are not just efficient but also effective in supporting livelihoods through promoting private sector participation.

3.6 Effectiveness

3.6.1 Definition and Intended Use

Effectiveness is defined as (Beck, 2006, p.49):

“Effectiveness measures the extent to which an activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criterion of effectiveness is timeliness.”

Beck (2006, pp.49-50) further explains the definition:

“Assessing effectiveness involves an analysis of the extent to which stated intervention objectives are met. […] Evaluation of effectiveness is therefore linked to evaluation of impact and longer-term effects of the intervention.”

The following issues to consider are highlighted in EHA publications:

- “Where possible, evaluations need to go beyond assessing activities and begin to examine who uses and benefits from resources provided, with data disaggregated by sex, socioeconomic grouping and ethnicity.” (Beck, 2006, p.50).
- Keeping a focus on why were interventions effective or not. (ibid.).
- “The evaluator’s role is to attempt to evaluate the intervention against stated objectives or agency or government standards, rather than activities. In particular, evaluators should ask whether and how primary stakeholders participated in the intervention’s design.” (Beck, 2006, p.51).
- Timeliness is a “key element in the assessment of effectiveness.”; “The phasing of interventions is often crucial to success, and evaluations should consider whether interventions were carried out in a fashion that adequately supported the affected population at different phases of the crisis.” (ibid.).

Following Beck (2006, p.54), the important aspects of *Coordination* (while not being a formal OECD-DAC criterion) will be considered under *Effectiveness* as coordination “cuts across several criteria” and these two are “closely related”. In brief, coordination is conceived as: *"The extent to which the interventions of different actors are harmonised with each other, to promote synergy, and to avoid gaps, duplication, and resource conflicts."* (Buchanan-Smith and Cosgrave, 2010?, p.39).

For coordination, the following issues to consider are highlighted in EHA publications:

- Focus on the role of the host government and other local institutions, and how coordination with non-traditional partners (e.g. military) is organised (Beck, 2006, p.55).
- "Given the multiplicity of actors involved in an emergency response, it is important that coordination is explicitly considered – the intervention of a single agency cannot be evaluated in isolation from what others are doing, particularly as what may be appropriate from the point of view of a single actor, may not be appropriate from the point of view of the system as a whole." (Hallam, 1998, p.56).

3.6.2 Applied Use, Key Findings and Further Considerations

The criterion effectiveness is used in nine out of ten evaluations. In the end, these had to measure the extent to which interventions have achieved hygiene improvements, i.e. the main purpose of any WASH emergency programme (*Figure 2*). In the evaluation reports however, this purpose is rarely clearly articulated as such under effectiveness. The reason for this is simply that evaluators can only evaluate what is actually measurable. Hence, unsurprisingly the subjects or issues used in conjunction with effectiveness primarily relate to outputs as subsequent *Table 7* shows. Next to coordination, technical aspects such as the design of WASH infrastructure or the choice of the HP approach are dominant themes addressed with this criterion. In two evaluations (Fortune and Rasal, 2010; Jeene, 2010) also the Sphere minimum standards in WASH are used as means for measuring a response's effectiveness. Particularly with respect to timeliness, the flexibility of an intervention or approach to adapt to changing or initially disregarded circumstances and needs is considered in several evaluation reports (DeVillez, Bousquet and Nyagwambo, 2011; Grayel and Mattson, 2012; Fortune and Rasal, 2010; Laurens, 2005).

Overall, the key findings from the evaluation reports (*Table 7*) demonstrate a high level of clarity in what is considered effective or ineffective. To some extent, the perception of effectiveness simply reflects the Sphere and HAP standards (e.g. the need for inclusive latrine design or participatory siting of WASH infrastructure). Another less surprising finding is that lack of monitoring with SMART (specific, measurable, attainable, relevant and time-bound) performance indicators necessarily limits the possibilities for evaluators to create evidence for which approaches work best in a given context. However, as e.g. the RTE of Fortune and Rasal (2010, pp.19-22) proves, simple observation techniques and interviews can already establish a good sense of if, for instance, the chosen HP method/approach is leading to any hygiene improvements, and what could be improved to be more effective in achieving this outcome. The great variety of issues represented in most key findings further highlight that the criterion effectiveness is best suited for weighing different aspects (which could severally be covered under other criteria) against each other.

Effectiveness	
	Subject/Issue
Criterion used in conjunction with:	The role of the local government and community in the response; Coordination with government, other NGOs, and through WASH committees and the WASH Cluster (Barham, Nabunny and Philpott, 2011a, pp.18-19); Coordination structure; Flexibility of the intervention approach; Standardised technical procedures (DeVillez, Bousquet and Nyagwambo, 2011, pp.15-17); Planning, monitoring and performance measurement tools; Coordination structures; Flexibility in response planning (Grayel and Mattson, 2012, pp.45-50); Cost-effectiveness; Inclusion of needs/capacities of beneficiaries (Fox, 2008, pp.17-19); Hygiene promotion approach; Flexibility in sanitation approach ;Sphere minimum standards in WASH; Coordination and cooperation mechanisms (Fortune and Rasal, 2010, pp.19-22); WASH infrastructure design; Cash transfer; Hygiene promotion approach (Luqman, Zulqarnain and Bahadar, 2012, p.40); WASH infrastructure design; Flexibility of project; Impact (Laurens, 2005, pp.23-40); Maintenance of WASH infrastructure; Timeliness of the response (Van der Wijk, Mwezi and Kazan, 2010, pp.17-19); Sphere minimum standards in WASH; Recruitment procedures; Security situation; Technical performance (Jeene, 2010, pp.16-17)
Selected key findings related to criterion:	Latrine design needs to be inclusive (e.g. to cater for persons with physical disabilities); It is vital to thoroughly think through ways of assisting people when phasing from free WASH services in camps to contributory services in areas of return to prevent any decline in WASH service functionality; Not including hygiene promotion – also in (donor) budgets – is ineffective; (Barham, Nabunny and Philpott, 2011a, pp.18-19); It is ineffective to not have standardised assessment tools and response procedures and not to comply with prescribed technical standards; Coordination is best done through a properly empowered body with executive powers – voluntary appointments restrict effectiveness; Only flexible intervention designs are effective (DeVillez, Bousquet and Nyagwambo, 2011, pp.15-17); It is very difficult to measure effectiveness without planning tools and SMART (specific, measurable, attainable, relevant and time-bound) performance criteria; Particularly large and long-lasting interventions need a strategic presence in the field with dedicated staff; Having best practices and lessons learned documentation to be shared with the WASH Cluster is vital (Grayel and Mattson, 2012, pp.45-50); The disadvantage of cost-effective implementation of interventions, e.g. installing several reverse osmosis water plants at the same time may, is likely to be a missing learning process with negative effects on sustainability (Fox, 2008, pp.17-19); Community mobilisation is of outmost importance to achieve good outcomes but requires the right approach and evidence based monitoring of its effectiveness (e.g. engaging HP volunteers to disseminate messages tent to tent may be effective if the style is not didactic in nature but is time consuming) (Fortune and Rasal, 2010, pp.19-22); Flexibility in latrine design is important to include e.g. special needs of physically challenged persons and avoid limited use; Short-term relief and recovery projects do not allow for enough time to set up effective community mobilisation mechanisms (Luqman, Zulqarnain and Bahadar, 2012, pp.17-19); Only participatory siting of WASH infrastructure is effective; Good quality workmanship needs time for training, follow-up technical controls, supervision and standard technical specifications (Laurens, 2005, pp.86-88); Not providing the systems, skills (trainings) and materials for long-term maintenance of WASH facilities is ineffective (Van der Wijk, Mwezi and Kazan, 2010, pp.17-19); The water quality should be tested for biological and chemical contamination before investing in rehabilitation of boreholes (Jeene, 2010, p.10)

Table 7: Effectiveness: Summary of use and key findings

This balancing of various relevant aspects is a key quality of effectiveness for evaluating WASH emergency response. For example Fox (2008, pp.17-19) exhibits that just emphasising cost-effective realisation of resource intensive interventions may not allow for a learning process during implementation to minimise negative effects when replicating the same on a large scale; hence, being possibly less effective in the end. Also timeliness matters, but the sometimes perceived need for speed can turn into a "counterproductive rush" with the effect of losing "opportunities for listening to people and building respectful relationships with local actors" and leading to "hasty decisions and inappropriate interventions" (Harvey, et al., 2010, p.31). However, as a survey among humanitarian practitioners revealed, poorly coordinated response efforts are seen as the single biggest problem or weakness hindering effectiveness (Taylor, et al., 2013, p.67). According to the meta-evaluation findings of Stokke (2007, p.16), the importance of coordination should also be stressed as a "prerequisite for sharing lessons learned between different agencies, as well as a mechanism for promoting accountability of humanitarian organisations".

In terms of its scope and depth in capturing essential processes of emergency response, effectiveness is the largest and most comprehensive OECD-DAC evaluation criteria. To address the identified WASH gaps, the effectiveness of many new and innovative WASH approaches need to be evaluated. E.g. in relation to hygiene promotion and community mobilisation in urban contexts, further development of "new tools and methodologies, like *sanitation marketing*" is necessary (Grünewald et al., 2011b, p.45). Lack of response monitoring based on SMART indicators certainly imposes a challenge for triangulating data and accordingly creating evidence. But as above considerations illustrate, and following Ramalingam, et al. (2009, p.85), the required mindset for evaluating effectiveness can be summarised with: *Not everything that can be counted counts, and not everything that counts can be counted*.

3.7 Impact

3.7.1 Definition and Intended Use

Impact is defined as (Beck, 2006, p.56):

"Impact looks at the wider effects of the project – social, economic, technical, environmental – on individuals, gender- and age-groups, communities and institutions. Impacts can be intended and unintended, positive and negative, macro (sector) and micro (household)."

Beck (ibid.) further explains the definition:

"Whereas assessment of effectiveness examines whether intermediate objectives have been achieved and fed into outcomes, assessment of impact usually examines the longer-term consequences of achieving or not achieving those objectives, and the issue of wider socioeconomic change. […] As such, assessment of impact often goes beyond intervention-planning documents, to consider the part the intervention plays in the wider socioeconomic and political contexts […] Because of their longer-term focus, evaluation of impact and connectedness are often closely linked."

The following issues to consider are highlighted in EHA publications:

- "Because of its wider scope, assessment of impact may not be relevant for all evaluations, particularly those carried out during or immediately after an intervention. […] evaluation of impact should be attempted only where: a longitudinal approach is being taken; there are data available to support longer-term analysis; the evaluation team includes specialists in socioeconomic and political analysis; and the commissioning agency is willing to invest in a more detailed evaluation." (Beck, 2006, pp.56-57).
- "Looking for impact could […] be misleading in the usually rather short timeframe of emergency humanitarian interventions. […] It may also become measureable on the very long term only, well after the end of humanitarian activities, and may not be attributable any more." (DG ECHO, 2007, p.34).
- "The further one moves from the time of the intervention, the more difficult it is to determine whether changes that have taken place are the result of the intervention or of some other factor, such as other interventions, or socioeconomic or political forces." (Beck, 2006, p.57).
- "Evaluation of impact needs to cover cross-cutting themes such as support to livelihoods, human rights, and gender equality." (ibid.).
- Impact evaluation "shifts the focus away from whether the anticipated amounts of aid were delivered or not (effectiveness) to whether people are better off or safer as a result" (Buchanan-Smith and Cosgrave, 2013, p.187).

3.7.2 Applied Use, Key Findings and Further Considerations

The criterion impact is used in six out of ten evaluations. All of those ultimately evaluated or rather tried to evaluate how effective interventions were in reducing or preventing WASH related diseases through hygiene improvements (*Figure 2*). But virtually none of the requirements listed for the criterion's intended use are applicable or existent in any of the evaluations. Therefore, impact assessment remains a noble idea whose aim and scope are well understood – as the subjects and issues addressed with this criterion (*Table 8* on the next page) demonstrate – but is extremely difficult up to infeasible to operationalise. The problems of attributing outcomes to specific intervention inputs are prevalent in all researched evaluations. Buchanan-Smith and Cosgrave (2013, p.187) point out that sole impact evaluations have been called a misnomer, as they "can more accurately be said to address outputs and outcomes". Similarly, the meta-evaluation of Taylor, et al. (2013, p.65) finds that in a significant number of cases, "the concept [of impact] is conflated with that of effectiveness". This altered use of impact is reflected well in the six evaluation reports and, for example, explicitly aimed for in the evaluation of Laurens (2005).

Regarding the definition of impact, the key findings included in *Table 8* can be summed up with two general findings. The first one is borrowed from the *ALNAP Annual Review* in 2003: "*Reports were so consistent in their criticism of agency monitoring and evaluation practices that a standard sentence could almost be inserted into all reports along the lines* *of: 'It was not possible to assess the impact of the intervention because of the lack of adequate indicators, clear objectives, baseline data and monitoring."* (ALNAP, 2003, p.107). The second one provides the evaluation of Fox (2008, p.24): *"In the meantime, it is safe to assume that the intervention will not have a negative impact on the health situation."*. The latter is as good as it gets when the impact of WASH interventions is evaluated. Besides effectiveness, all other findings (e.g. the serious health and environmental consequences when discharging untreated wastewater from built latrines to open drains or irrigation channels) are better placed under other evaluation criteria, notably connectedness or can be encompassed with cross-cutting issues as well as the Sphere and HAP standards.

Impact	
	Subject/Issue
Criterion used in conjunction with:	Vulnerable groups; WASH service provision; Communicable diseases (Barham, Nabunny and Philpott, 2011a, pp.19-20); Delivering of basic WASH services; Controlling high-risk waterborne diseases (Grayel and Mattson, 2012, p.48); Planning and Implementation of the intervention; Public health (Fox, 2008, pp.11-14); Environment; Sanitation-related diseases; Efficiency of response design (Fortune and Rasal, 2010, pp.8-30); Lives of the affected people; WASH behaviours and practices; Environment (Luqman, Zulqarnain and Bahadar, 2012, pp.35-39); WASH service provision; Sphere minimum standards in WASH (Jeene, 2010, p.18)
Selected key findings related to criterion:	No noticeable impact on reduction in communicable diseases such as diarrhoea as well as behavioural changes in basic hygiene, sanitation practices and safe water chain (Barham, Nabunny and Philpott, 2011a, pp.19-20); The intervention brought a minimum level of control for high-risk waterborne diseases and low humanitarian basic WASH services; Given the lack of outcome data, it is not possible to show what has changed (e.g. improved hygiene behaviours) (Grayel and Mattson, 2012, p.48); Because no baseline study was conducted prior to the intervention, it is only safe to assume that the intervention will not have a negative health impact (Fox, 2008, pp.24-25); For mass sanitation interventions in an urban context, there is an increased need to consider environmental impact and decommissioning of facilities to not compromise any future use of the land on which latrines have been built (Fortune and Rasal, 2010, p.8); "The discharge of untreated wastewater from the septic tank of latrines to open drains or irrigation channels has serious health and environmental consequences" (Luqman, Zulqarnain and Bahadar, 2012, p.35); While the remaining host population will have increased access to water, those who return to their places of origin will fall back to unprotected sources of water (Jeene, 2010, p.18)

Table 8: Impact: Summary of use and key findings

Despite an increasing interest in assessing the impact of humanitarian response through EHA, guidance on impact evaluation (e.g. Rogers, 2012) suggest that evaluating impact is an issue on its own that requires a quite different methodological foundation than current EHA has. One of the most frequently discussed methods necessary to establish evidence that a specific intervention will have a desired impact is *randomised controlled trial (RCT)*. So far, this method provides the most rigorous way of determining whether a cause-effect relation exist. RCTs are expensive, hence rarely used and also raise ethical concerns as they mean comparing a disaster-affected population group receiving humanitarian assistance with one that does not. In response to such difficulties and concerns, new evaluation concepts are being developed for humanitarian contexts. One of the most recent examples published is *Contribution to Change: An approach to evaluating the role of intervention in disaster recovery* (Few, et al., 2014)[4]. But also this approach requires an evaluation design with an exclusive focus which excludes the scope of the other OECD-DAC criteria.

With Lindahl (1998, p.9) it can be concluded, that "[i]mpact assessment, as well as measuring efficiency and effectiveness is not an *ad hoc* activity that an evaluation can carry out **unless** the programme has paid considerable attention to this already from its inception". The "future of humanitarian impact assessment", Proudlock et al. (2009, p.74) argue, "lies in linking different partners across the sector, and introducing new actors and perspectives". The authors note that "long term partnerships between donors and recipient governments, implementing agencies, academics and recipient communities are key to ensuring sector-wide and institutional sustainability of impact assessment" (ibid.). The humanitarian system has made important steps in this regard – particularly through the cluster system (with the WASH clusters being among the most active and innovative ones) – but its current state (Harvey, et al., 2010; Taylor, et al., 2013) does not suggest that impact evaluations could bring about any significant changes that the commitment to quality and accountability standards could not have already accomplished over the past decades.

3.8 Integration of Cross-cutting Themes

The analysis of the OECD-DAC evaluation criteria shows that under which criterion findings are placed is rather a matter of convenience. Generally speaking, the same is true for integrating cross-cutting themes with the OECD-DAC evaluation criteria. Following Beck (2006), the only thing that matters is that a "comprehensive evaluation is carried out, using all of the criteria where relevant" (p.19) and when applying them, to cover key elements of important EHA themes (p.78). Cross-cutting themes are used in all researched evaluations. As *Table 9* on the page following the next summarises, the themes emphasised vary, and depending on the context of the emergency intervention as well as the thematic focus of an evaluation, some may be more relevant than others. In the ALNAP Quality Proforma (*Annex 1*), the Sphere and/or HAP standards are viewed as cross-cutting themes too. To use standards as such rather than equivalent criteria based on the standards' intention is apparent in the majority of evaluations. In fact, there are considerable overlaps between these themes and humanitarian standards, and the Sphere handbook too outlines a set of cross-cutting themes to be considered when applying the Sphere standards (see The Sphere Project, 2011, pp.14-17).

A certain level of vagueness and inconsistency in the use of cross-cutting themes and standards together as quality aspects to enhance the OECD-DAC evaluation criteria can be

[4] As the title says, this guide considers the contribution to outcomes or contributory impact. The approach is designed for situations where there has been an external intervention following rapid-onset natural hazards and intended to foster people's recovery in the medium term. The described method focuses on assessing positive and negative changes to the lives of affected people and other local stakeholders and assumes that changes in people's well-being and livelihoods can be most clearly identified at the household level (Few, et al., 2014, p.7).

attested from the sample of evaluation reports. The cross-cutting themes themselves are mere generic expressions of fundamental concerns in disaster response. They are typically phrased in the same manner in the evaluation questions of the selected evaluations, meaning that how extensively they are addressed is up to the evaluator's understanding of the context. The OECD-DAC framework of Beck (2006) suggest the following list of cross-cutting themes for EHA (for the succinct explanations see pp.78-80):

- *The influence and understanding of local context*
- *Human resources and management*
- *Protection*
- *Participation of primary stakeholders*
- *Coping strategies and resilience*
- *Gender equality*
- *Environment*

With the ALNAP Quality Proforma and the Sphere handbook, the following additions can be made:

- *Capacity building*
- *DRR*
- *Advocacy*
- *Vulnerable and marginalised groups (e.g. children, older people, people living with HIV/AIDS or persons with disabilities)*
- *(Psychosocial support)*

This listing is essentially another compilation of weak spots in humanitarian action and except for psychosocial support, all of the cross-cutting themes directly relate to WASH. For the most part, these are covered by the Sphere and HAP standards and in particular Sphere offers a more nuanced understanding of what should be expected from emergency response in relation to these issues. For instance Sphere's *Protection Principles*[5], which inform the Sphere core and minimum standards, provide substantial guidance on protection as well as its limits. But regardless of elaborated and comprehensive humanitarian standards, cross-cutting themes can serve as important headings in evaluations to provide a more (quality) focused analysis (e.g. if an intervention has also been effective in providing accessible WASH services to identified vulnerable groups). However, to gain analytical depth, using standards is inevitable, not least because they feature reasonable indicators.

[5] The four protection principles – described in detail in the Sphere handbook (The Sphere Project, 2011, pp.25-47) – are:

- *Avoid exposing people to further harm as a result of your actions*
- *Ensure people's access to impartial assistance – in proportion to need and without discrimination*
- *Protect people from physical and psychological harm arising from violence and coercion*
- *Assist people to claim their rights, access available remedies and recover from the effects of abuse*

Cross-cutting Themes	
	Subject/Issue
Themes used (and with which criterion):	Environment; Gender combined with protection (Grünewald, et al., 2011a, p.34); Gender/Participation of women (*Sustainability*); Vulnerable and marginalised groups and Sphere minimum standards in WASH (*Effectiveness*) (Barham, Nabunny and Philpott, 2011a, pp.19-20); Environmental protection; Gender; HIV/AIDS; Children/Girls; Ownership/Participation (DeVillez, Bousquet and Nyagwambo, 2011, pp.24-26); Beneficiary accountability/participation; Vulnerable groups; Humanitarian standards (dignity, gender considerations, protection, assistance to displaced persons, disaster resilience, etc.) and Sphere minimum standards in WASH (Grayel and Mattson, 2012, pp.20-47); Ownership/Participation and gender (*Sustainability*); DRR and capacity building (Fox, 2008, pp.20-21); Environment; Gender; Accountability/Participation; Sphere minimum standards in WASH; Vulnerable and marginalised groups (Fortune and Rasal, 2010, pp.9-30); Gender and protection; DRR and capacity building; Participation of affected people; Environment; Sphere minimum standards in WASH; Accountability and complaint mechanism (HAP standard) (Luqman, Zulqarnain and Bahadar, 2012, pp.17-40); Gender; DRR; Capacity building; Participation/Ownership (Laurens, 2005, pp.23-74); Protection; Gender; Adherence to humanitarian standards (Code of Conduct, Sphere and HAP) (*Effectiveness*); Vulnerable and marginalised groups (*Coverage*) (Van der Wijk, Mwezi and Kazan, 2010, pp.14-34); Gender and protection; (*Relevance/Appropriateness*) Sphere minimum standards in WASH (*Coverage*, *Effectiveness*, *Impact*); Advocacy (Jeene, 2010, pp.7-20)
Selected key findings related to themes:	Limited integration of gender issues (e.g. design and siting of latrines and showers) meant that people, and women in particular, were at real risk in certain camps and neighbourhoods (Grünewald, et al., 2011a, p.34); "All the institutional sanitation structures (latrines) were not designed to cater for persons with physical disabilities" (Barham, Nabunny and Philpott, 2011a, p.19); During times of urgent needs for relief supplies in the immediate response phase it may be difficult, even infeasible, to emphasise common cross-cutting issues (DeVillez, Bousquet and Nyagwambo, 2011, pp.24-26); It is vital that WASH programmes incorporate strong community mobilisation components to monitor barriers experienced by vulnerable populations and the identification of solutions to address barriers (Grayel and Mattson, 2012, pp.21-22); "the women had asked for household taps, but been informed that this was not part of the project. Perhaps it would have been wise at this point to reconsider the type of project and its aim since it is clearly beyond emergency water supply" (Fox, 2008, p.21); In countries with already high levels of deforestation, not sourcing wood internationally (e.g. for large scale latrine construction programmes) contributes to environmental degradation and future disasters (Fortune and Rasal, 2010, p.24); "In remote […] water scarce areas where women fetch water from distances taking hours, the construction of water-based latrine[s] should be discouraged. The focus in such areas should be on water supply followed by latrines." (Luqman, Zulqarnain and Bahadar, 2012, p.35); "women complain that they have not been consulted about suitable locations of the water points" (Laurens, 2005, p.74); "WASH facilities are a source of protection for the most vulnerable beneficiaries, particularly women and children." (Van der Wijk, Mwezi and Kazan, 2010, p.34); "Gender and protection must be part of hygiene promotion in displacement." (Jeene, 2010, p.20)

Table 9: Cross-cutting themes: Summary of use and key findings

As demonstrated earlier, the *influence and understanding of the local context* is an underlying theme in evaluations when considering relevance/appropriateness respectively effectiveness; likewise, *human resources and management* are part of efficiency. All other addressed cross-cutting themes – and which are less apparent from the OECD-DAC criteria definitions – are shown in *Table 9*. According to the analysis, four broad thematic areas can be differentiated:

- Gender and protection
- Environment
- Resilience, DRR and capacity building
- Accountability, participation, the inclusion of marginalised/vulnerable groups and similar themes which express key contents of the Sphere and HAP standards

Although these themes are only partially considered directly with the OECD-DAC criteria (*Table 9*), *connectedness and effectiveness* are doubtless core evaluation criteria to capture several of them. But while it may be obvious to address an intervention's considerations given to environment, resilience, DRR and capacity building with connectedness, those given to the remaining themes do not always fit best with the intentions of other evaluation criteria. In several instances (especially where the OECD-DAC criteria are also used to structure findings) cross-cutting themes remain just mentioned (see e.g. the evaluation of Jeene (2010)). Following ECHO's proposal of small set of generic evaluation criteria that includes humanitarian *values* as a separate criterion (*Chapter 2*), it appears worth considering the standards relevant to WASH at the basis of a quality criterion (including accountability) in addition to a limited set of OECD-DAC criteria. As evident from the evaluation reports, evaluators typically face several cross-cuttings themes at a time when assessing overall performance of WASH interventions. Thus, the Sphere and HAP standards can be suggested as a practical tool to maintain a focused evaluation on quality, at the same time ensuring that major issues aren't neglected.

All of the key findings from the researched evaluations (*Table 9*) can be regarded valuable for learning albeit some are not news. They articulate key issues that make a difference in the lives of disaster-affected people and regrettably point out the relevance of EHA. So far, the analysis in this chapter has shown that what can be considered with cross-cutting themes adds important quality aspects to the scope of the OECD-DAC criteria and is vital to assessing general performance of emergency response. Because several issues conceived with cross-cutting themes are rather related to the entire intervention than single performance aspects expressed through the OECD-DAC criteria, and simply because standards convey what humanitarian response strives to achieve, employing a corresponding quality criterion in evaluations is reasonable. To make way for it and as the preceding examination reveals, the OECD-DAC evaluation criteria can be reduced to a few relevant ones, i.e. *connectedness, efficiency and effectiveness* (to be seen however in the emergency contexts of the researched evaluations).

4 Using Quality and Accountability Standards for Evaluations in the WASH Sector

Following up on the meta-analysis of evaluation reports, this chapter finally explores how quality and accountability standards are used in evaluations. The first section presents major and comprehensive general standards, the *Sphere Core Standards* and *HAP Standard Benchmarks*. Based on insights from their applied use and findings from *Chapter 3*, lessons for EHA will be drawn. The second section then deals with WASH specific standards, the *Sphere Minimum Standards in Water Supply, Sanitation and Hygiene Promotion*, and similarly analyses theory and praxis to elaborate on identified gaps. As in the previous part of the analysis, an attempt has been made to retrieve learning-focused findings from the evaluation reports. These are therefore referred to as selected key findings in *Table 10 and 11*.

4.1 General Standards in Humanitarian Response

4.1.1 Definition of the Sphere Core Standards and HAP Standard Benchmarks

The Sphere core standards and the HAP standard benchmarks contain requirements that complement each other to promote “coherence” and reinforce a “shared commitment to accountability” (The Sphere Project, 2011, p.53; HAP International, 2013, p.54). The definition of the six Sphere core standards are displayed in *Figure 10* on the next page – the full standard with key actions, key indicators and guidance notes is published in the *Humanitarian Charter and Minimum Standards in Humanitarian Response* (The Sphere Project, 2011, pp.49-78). Subsequently, *Figure 11* depicts the definitions of the six HAP standard benchmarks – the full standard with all specific requirements is published in the *Guide to the 2010 HAP Standard in Accountability and Quality Management* (HAP International, 2013, pp.23-51). In the following, a linking of these twelve standards based on their complementary aspects outlines key points accompanying the definitions, based on the guidance notes (The Sphere Project, 2011, pp.49-78) and requirements (HAP International, 2013, pp.23-51) for each standard/standard benchmark:

- **People-centred humanitarian response/Participation/Handling complaints**

  Support and build on local capacity; Provide detailed mechanisms for: Feedback, complaints, redress and influencing decision making (informed consent as a minimum), assessment, project design/implementation and M&E; Ensure representative participation

- **Coordination and collaboration/Sharing information**

  Identify local civil society actors and networks involved in the response; On principle it is the affected state’s role to coordinate humanitarian response; Share information in appropriate/accessible languages, formats and media

- **Assessment**

  Ensure collaborative pooling of pre-disaster information; Assessment is a process, not a single event; Disaggregate data by sex and age; Assess vulnerability and community capacities for coping and recovery; Ensure representative assessments

- **Design and response/Establishing and delivering on commitments**

  Support existing capacity; Ensure people's dignity, access to aid (non-discrimination) and context, vulnerability and conflict sensitivity; Meet Sphere's minimum standards and where/when possible increase the scope of the response to include the host population; Take early recovery and risk reduction actions at the earliest opportunity; Set out an accountability framework and a corresponding quality management system

- **Performance, transparency and learning/Learning and continual improvement**

  Monitor impact, relevance and performance (according to Sphere and HAP standards)

- **Aid worker performance/Staff competency**

  Consistently review/act on staff performance (e.g. skills, behaviour); Transparent recruitment


Core Standard 1: People-centred humanitarian response

People's capacity and strategies to survive with dignity are integral to the design and approach of humanitarian response.

Core Standard 2: Coordination and collaboration

Humanitarian response is planned and implemented in coordination with the relevant authorities, humanitarian agencies and civil society organisations engaged in impartial humanitarian action, working together for maximum efficiency, coverage and effectiveness.

Core Standard 3: Assessment

The priority needs of the disaster-affected population are identified through a systematic assessment of the context, risks to life with dignity and the capacity of the affected people and relevant authorities to respond.

Core Standard 4: Design and response

The humanitarian response meets the assessed needs of the disaster-affected population in relation to context, the risks faced and the capacity of the affected people and state to cope and recover.

Core Standard 5: Performance, transparency and learning

The performance of humanitarian agencies is continually examined and communicated to stakeholders; projects are adapted in response to performance.

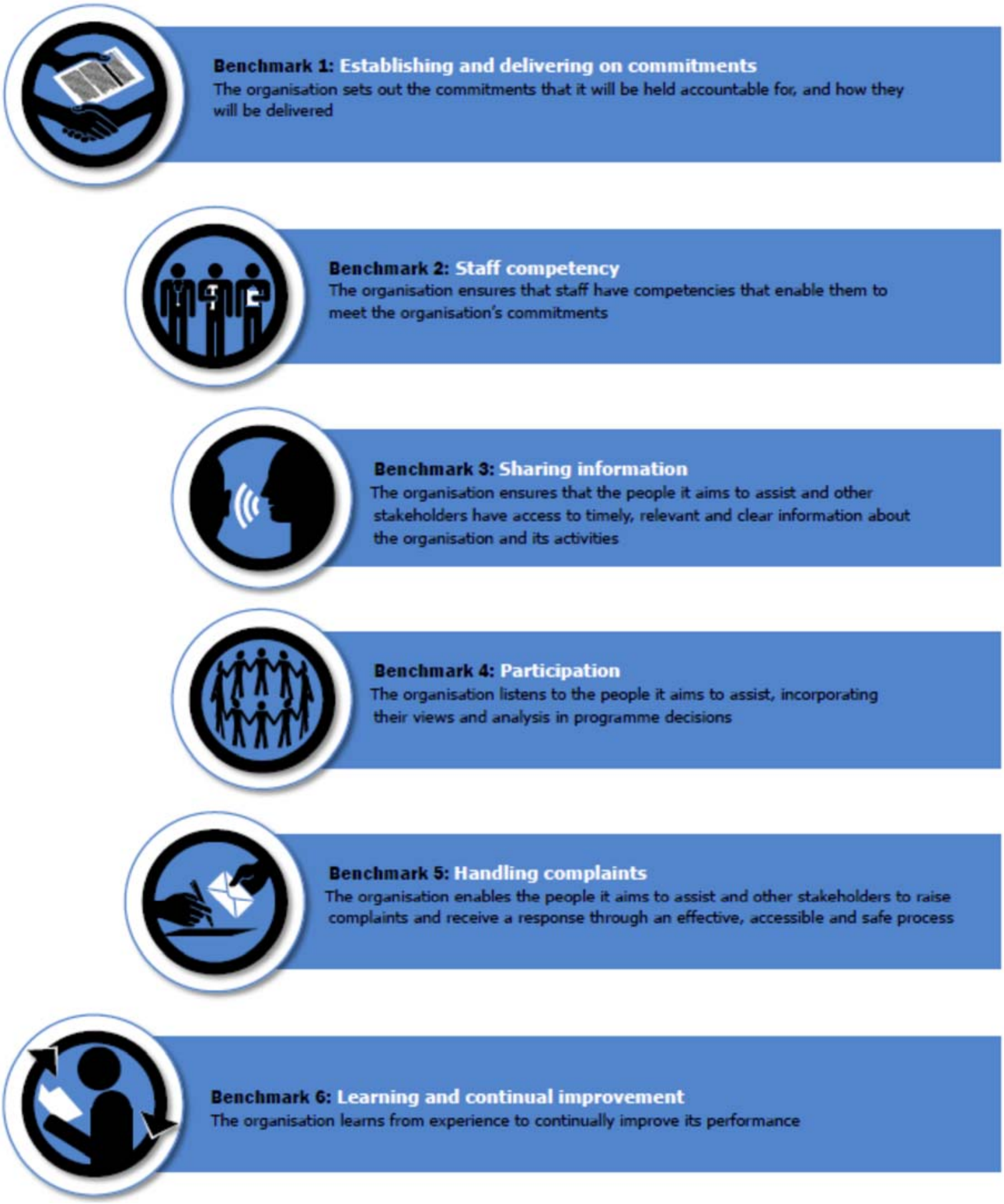
Source: The Sphere Project, 2011, pp.55-73

Figure 10: The Sphere core standards

Core Standard 6: Aid worker performance

Humanitarian agencies provide appropriate management, supervisory and psychosocial support, enabling aid workers to have the knowledge, skills, behaviour and attitudes to plan and implement an effective humanitarian response with humanity and respect.

Figure 10 (continued): The Sphere core standards



Source: HAP International, 2013, p.21

Figure 11: The HAP standard benchmarks

4.1.2 Applied Use, Key Findings and Further Considerations

As outlined in *Figure 10 and 11*, the Sphere core standards and HAP standard benchmarks contain elements relevant to any humanitarian response regardless of the type of emergency. Selected Sphere core standards or HAP standard benchmarks are referenced in all researched evaluations; five evaluations are referring to all Sphere core standards and HAP standard benchmarks (*Table 10* below). So, overall, it can be said that the importance of the issues addressed in these general standards is well understood by evaluators. But as with any evaluation criteria, how they are used matters much more for the quality of an evaluation than if they are used. In available EHA guides humanitarian standards play a minor role compared to the OECD-DAC criteria. They appear rather as a side note and there is no guidance on how to work with them. Following Beck (2006, p.19), regardless of the kind of evaluation criteria used, “evaluators should ask both what happened, and why it happened”.

Perhaps, evaluating against humanitarian standards invites much more to focus on *what* questions than *why* questions. At least, this is what many examples in the researched evaluation reports suggest. This practice runs counter to lesson-learning. Except for coordination (admittedly less related to the Sphere core standard *coordination and collaboration*), in the majority of evaluations, specific core standards or standard benchmarks are not explicitly addressed – rather certain aspects covered by them are referenced. Only two evaluations are different in this regard, with one (Van der Wijk, Mwezi and Kazan, 2010) assessing separately accountability and quality management, however in a sketchy manner, and with another one (Luqman, Zulqarnain and Bahadar, 2012) elaborating findings on the implemented complaints handling mechanism.

While it is negligible if standards are explicitly referenced, but to make sure that their core intent is considered, regularly a superficial and/or inconsistent use (similar to what the analysis of cross-cutting themes revealed) can be observed in the selected evaluations. The choice of sound and qualified findings to be included in *Table 10* was small and in some cases barely a choice. Another telling insight from the meta-analysis is that although the Sphere standards are much more often explicitly addressed in evaluation reports than the HAP standard, what is usually meant with Sphere standards, is the Sphere minimum standards in WASH. There is clearly a limited understanding of the Sphere concept (as explained at the end of *Chapter 2*) by many evaluators. It is ironic that standards such long in existence and common to humanitarian response are not properly grasped.

The key findings from the evaluation reports presented below in *Table 10* once more point out the importance of the Sphere core standards and complementary HAP standard benchmarks. The issues addressed are rather down-to-earth, but being as simple as effective in impacting the success of WASH emergency response. Because none of the quality and accountability standards is a how-to guide, and their use is always more straightforward on paper than in the real world, EHA could do well in gaining better insights into the factors affecting their implementation. For instance, constitutive to accountability is handling complaints, but corresponding procedures will have to be different according to the context, also culturally. EHA can be used to create evidence on what worked out in which situation.

	Sphere Core Standards and HAP Standard Benchmarks
Standards/ references to standards used:	Assessment (Grünewald, et al., 2011a, p.33); People-centred humanitarian response/Participation; Coordination and collaboration (Barham, Nabunny and Philpott, 2011a, pp.17-21); People-centred humanitarian response/Participation; Coordination and collaboration; Aid worker performance/Staff competency (DeVillez, Bousquet and Nyagwambo, 2011, pp.11-26); All Sphere core standards/HAP standard benchmarks (Grayel and Mattson, 2012, pp.15-47); People-centred humanitarian response/Participation; Coordination and collaboration/Sharing information; Assessment; Design and response/Establishing and delivering on commitments; Performance, transparency and learning/Learning and continual improvement; Aid worker performance/Staff competency (Fox, 2008, pp.11-23); All Sphere core standards/HAP standard benchmarks (Fortune and Rasal, 2010, pp.10-29); All Sphere core standards/HAP standard benchmarks (Luqman, Zulqarnain and Bahadar, 2012, pp.15-32); All Sphere core standards/HAP standard benchmarks (Laurens, 2005, pp.23-74); All Sphere core standards/HAP standard benchmarks (Van der Wijk, Mwezi and Kazan, 2010, pp.14-33); People-centred humanitarian response/Participation; Coordination and collaboration; Design and response; Aid worker performance/Staff competency (Jeene, 2010, pp.7-18)
Selected key findings related to standards:	"initial assessments which preceded water and sanitation programmes did not look into pre-existing practices in terms of access to water and excreta management"; It took a long time until e.g. the existence of standpipes and water provider networks were taken into account in organisations' strategies (Grünewald, et al., 2011a, p.33); The move from the national WASH cluster to local WASH committees has not been very effective (Barham, Nabunny and Philpott, 2011a, p.19); "many field staff were quite junior in the INGO hierarchy and did not have much in-house expertise. This necessarily inhibited their effectiveness in performing their duties as they were receiving on the job training in a learning-by-doing mode." (DeVillez, Bousquet and Nyagwambo, 2011, p.19); "while two-way communication is emphasized the identified two-way methods recommended (talk back radio, television productions that interact with beneficiaries in the field, publications with beneficiary news and use of SMS to collect feedback on specific questions/issues) provide limited capacity for engaging beneficiaries in a real meaningful two-way dialog" (Grayel and Mattson, 2012, p.35); The government suggested to provide water desalination units to 20 islands. "This was initially accepted without apparent follow-up with baseline assessment. This has made it difficult to generate adequate insight into the actual needs for these systems in particular in regards to alternative water supply systems on the islands." (Fox, 2008, p.11); "With information, people are able to choose whether or not to engage. […] people spoken to in the camps did not seem to know who was responsible for the sanitation services, including who was cleaning them. The result of this is a reduced likelihood that people will contribute to the development, implementation or monitoring of the facilities." (Fortune and Rasal, 2010, p.27); "There is a clear need […] to invest more on integration of Humanitarian Accountability principles across […] programs and initiate capacity building […] in HAP." (Luqman, Zulqarnain and Bahadar, 2012, p.32); Local perception of the founded community water committees/cash box collection system: "systematic lack of transparency in respect of financial issues" (Laurens, 2005, p.81); More training of local partners is needed to improve the adherence to humanitarian standards and fundamental principles (Van der Wijk, Mwezi and Kazan, 2010, p.33); Slow recruitment was a considerable delay factor (Jeene, 2010, pp.16-17)

Table 10: Sphere core standards and HAP standard benchmarks: Summary of use and key findings

In this sense, EHA is also a tool to identify best practices which can then be elaborated and published with more suitable methods, e.g. case studies. The mainstreaming of cash-based programming noted earlier in the analysis of efficiency illustrates this potential for innovations in humanitarian response. The Sphere and HAP standards have a strong foundation grounded in practical evidence on quality and accountability. Hence, evaluations should build on it not rephrase it to provide informative findings, at the same time questioning them and pinpointing areas of improvement. If one compares the wealth of relevant information provided by Sphere and HAP with key findings from the evaluation reports, it must be acknowledged that the additional value of the latter is marginal in many areas the standards address. Anyhow, humanitarian standards do not provide answers to anything and are limited in some major cross-cutting areas, notably resilience and DRR. As e.g. Stokke (2007, p.18) emphasises, there is a "need to go from humanitarian response to proactive preparedness and vulnerability reduction". Where this is a focus in EHA, the increasing amount of existing research and guidance[6] for humanitarian action in such areas should be considered to provide more realistic, specific and relevant recommendations to decision makers (for the quality of recommendations concerning WASH and DRR see e.g. Luqman, Zulqarnain and Bahadar, 2012, pp.34-35).

Not least because EHA is a time-limited activity under resource constraints as outlined in *Chapter 2*, focussing on a few core standards/standards benchmarks can yield better results than the observed evaluation practice of having a rather broad but cursory scope. Since several of them relate in one way or the other to single OECD-DAC criteria (see the outlined key points accompanying the Sphere and HAP definitions above), some standards/standard benchmarks are ideally conceived together with the chosen OECD-DAC criteria. E.g. the Sphere core standard/HAP standard benchmark *design and response/establishing and delivering on commitments* can be considered under *effectiveness*. Moreover, as *Chapter 3* demonstrated, there is nothing inherent in the OECD-DAC evaluation criteria which stops evaluators from enhancing them through quality and accountability dimensions of the Sphere and HAP standards.

However, for practical reasons it might be sometimes more suitable to specify selected standards/standard benchmarks in an additional quality evaluation criterion. To conclude, the general nature of the Sphere core standard and HAP standard benchmark definitions should not detract from the accompanying key actions, key indicators, guidance notes or requirements as (so far) the only available means of measuring general quality and accountability in humanitarian action. Particularly for M&E, the GWC has drafted an accountability checklist for WASH based on the HAP standard (*Annex 2*). One could think of further refining that list and using it at the basis of a strategically selected evaluation with the purpose to just learn about how accountability can be achieved in different phases of emergency response and how specific accountability measures impact on the quality of response. Such an evaluation would demonstrate a much closer understanding of accountability than classic accountability-oriented evaluations required for donor reporting.

[6] See e.g. *Toward Resilience. A Guide to Disaster Risk Reduction and Climate Change Adaption* (Turnbull, Sterrett and Hilleboe, 2013) with a separate section on WASH, and the publication's listed tools and resources.

Evaluation is a tool, and as it is with any tool, its potentials are only limited by the imagination of the user (i.e. humanitarian agencies) than the tool itself.

4.2 WASH Specific Standards in Humanitarian Response

4.2.1 Definition of the Sphere Minimum Standards in WASH

The Sphere minimum standards in WASH are the most comprehensive and elaborated ones specifically addressing WASH issues in humanitarian response, apart from possible national regulations. *Figure 12* displays the definitions of the 13 minimum standards covering the areas *programme design/implementation, hygiene promotion, water supply, excreta disposal, vector control, solid waste management and drainage*. The full standard detailing its application with key actions, key indicators and guidance notes is published in the *Humanitarian Charter and Minimum Standards in Humanitarian Response* (The Sphere Project, 2011, pp.79-137). The Sphere Project (2011, p.80) emphasises that the *Protection Principles* and *Core Standards* "must be used consistently" with the minimum standards, which may also be considered "during disaster preparedness and the transition to recovery activities". It is important to keep in mind that the minimum standards are "qualitative in nature and specify the minimum levels to be attained" (ibid.). In the following, key points in attaining the minimum standards are listed for each area, based on their guidance notes (The Sphere Project, 2011, pp.49-78):

- **Programme design and implementation**

  An assessment is needed to identify risky practices, resources available to the population, local knowledge/practices and social/cultural norms in order to make WASH facilities and HP approaches effective, relevant and successful

- **Hygiene promotion**

  Ensure that no group (e.g. women) is overburdened with responsibility for HP activities – consider benefits (e.g. training, employment opportunities); Realise that e.g. privacy, safety, convenience, social status or esteem may be more important motivators for behavioural change than health; Get people's priorities right (e.g. distributed hygiene items may be sold if livelihood priorities area not yet met)

- **Water supply**

  The quantities of water needed depend on the context; Pay close attention to post-delivery contamination

- **Excreta disposal**

  Consider the issues of toilets in difficult environments (e.g. final disposal of sludge, collection and disposal system of plastic bags containing faeces) and their implications for the HP approach

- **Vector control**

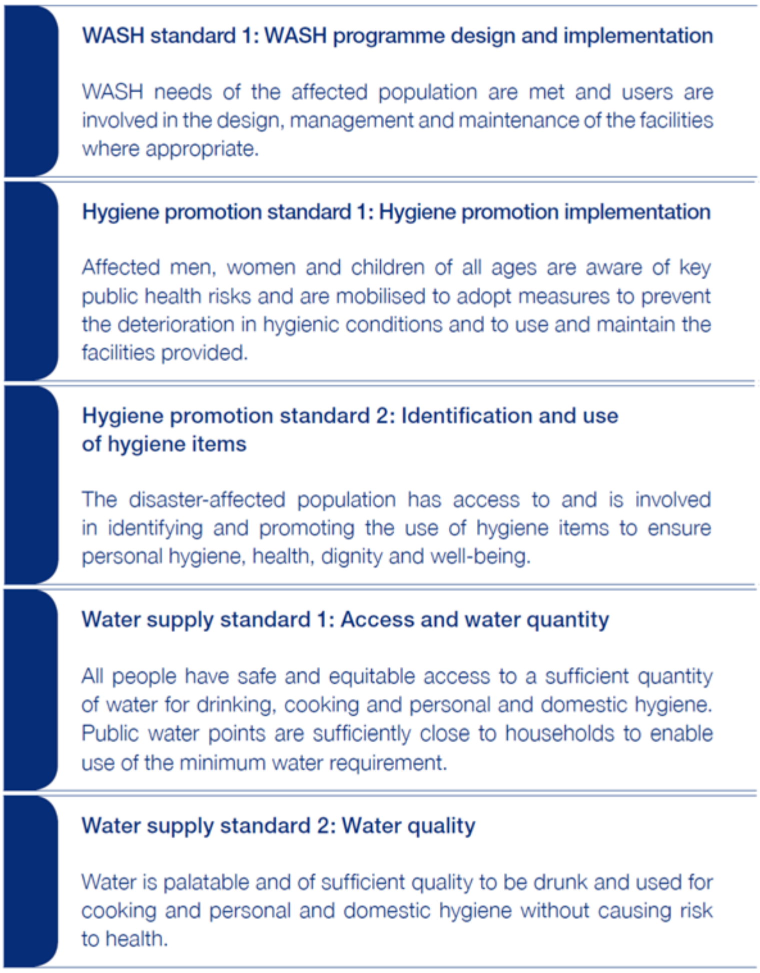
  Vector control is a delicate issue, involves clinical evidence of a vector-borne disease problem and adherence to international protocols and norms; interventions "may have no impact on disease if they target the wrong vector, use ineffective methods or target the right vector in the wrong place or at the wrong time"

- **Solid waste management**

  Requires to be planned and implemented in close consultation with the affected population and relevant agencies/authorities

- **Drainage**

  Most effectively controlled through site selection; Ensure on-site disposal where possible; Involve the affected population and use local knowledge of natural flow of drainage water



WASH standard 1: WASH programme design and implementation

WASH needs of the affected population are met and users are involved in the design, management and maintenance of the facilities where appropriate.

Hygiene promotion standard 1: Hygiene promotion implementation

Affected men, women and children of all ages are aware of key public health risks and are mobilised to adopt measures to prevent the deterioration in hygienic conditions and to use and maintain the facilities provided.

Hygiene promotion standard 2: Identification and use of hygiene items

The disaster-affected population has access to and is involved in identifying and promoting the use of hygiene items to ensure personal hygiene, health, dignity and well-being.

Water supply standard 1: Access and water quantity

All people have safe and equitable access to a sufficient quantity of water for drinking, cooking and personal and domestic hygiene. Public water points are sufficiently close to households to enable use of the minimum water requirement.

Water supply standard 2: Water quality

Water is palatable and of sufficient quality to be drunk and used for cooking and personal and domestic hygiene without causing risk to health.

Source: The Sphere Project, 2011, pp.88-123

Figure 12: The Sphere minimum standards in WASH

Water supply standard 3: Water facilities

People have adequate facilities to collect, store and use sufficient quantities of water for drinking, cooking and personal hygiene, and to ensure that drinking water remains safe until it is consumed.

Excreta disposal standard 1: Environment free from human faeces

The living environment in general and specifically the habitat, food production areas, public centres and surroundings of drinking water sources are free from human faecal contamination.

Excreta disposal standard 2: Appropriate and adequate toilet facilities

People have adequate, appropriate and acceptable toilet facilities, sufficiently close to their dwellings, to allow rapid, safe and secure access at all times, day and night.

Vector control standard 1: Individual and family protection

All disaster-affected people have the knowledge and the means to protect themselves from disease and nuisance vectors that are likely to cause a significant risk to health or well-being.

Vector control standard 2: Physical, environmental and chemical protection measures

The environment where the disaster-affected people are placed does not expose them to disease-causing and nuisance vectors, and those vectors are kept to a reduced level where possible.

Vector control standard 3: Chemical control safety

Chemical vector control measures are carried out in a manner that ensures that staff, the disaster-affected population and the local environment are adequately protected and that avoids creating chemical resistance to the substances used.

Solid waste management standard 1: Collection and disposal

The affected population has an environment not littered by solid waste, including medical waste, and has the means to dispose of their domestic waste conveniently and effectively.

Drainage standard 1: Drainage work

People have an environment in which health risks and other risks posed by water erosion and standing water, including stormwater, floodwater, domestic wastewater and wastewater from medical facilities, are minimised.

Figure 12 (continued): The Sphere minimum standards in WASH

4.2.2 Applied Use, Key Findings and Further Considerations

Selected Sphere minimum standards in WASH are used in all evaluations (*Table 11*). Or rather their use is indicated, because what is in many instances used is not the standard itself but figures provided as part of the key indicators or guidance notes. A classic example of such a figure often referred to is 15 litres of potable water per person per day as an indicator for meeting basic survival needs. To further illustrate this example, the corresponding guidance note says: *"The quantities of water needed for domestic use is context based, and may vary according to the climate, the sanitation facilities available, people's habits, their religious and cultural practices, the food they cook, the clothes they wear, and so on. Water consumption generally increases the nearer the water source is to the dwelling. Where possible, 15 litres per person per day (l/p/d) can be exceeded to conform to local standards where that standard is higher."* (The Sphere Project, 2011, p.98). The standard referred to – *Water supply standard 1: Access and water quantity* (*Figure 12*) – is however on purpose and necessarily a qualitative one. Furthermore, it addresses other aspects than water quantity and those are again explained in more detail in the respective guidance notes.

This short examples demonstrates an often observed mindset when using the Sphere minimum standards in WASH in evaluations: Using isolated aspects of a standard without any or little relation to the standard. This is what the research does definitely not suggest when arguing that e.g. key indicators or guidance notes contain valuable information and should be used as means of measuring quality. The problem is not about weather specific figures are used or not, but that evaluators (next to many humanitarian staff) tend to reduce the Sphere standards to something quite different than what they are – often turning out to be mere a couple of numbers that have little to do with quality, and accountability too. Such misconceptions then can lead up to following misguided declarations in evaluation reports: *"SPHERE minimum standards may not be relevant here since the affected population are living in their homes. Most of them have some access to water quantity which may not have safe quality. Similarly, in the selected villages the concept of community latrine is not applicable and people do not use neighbours latrines so the distance to latrines and number of people per latrines are not relevant as proposed by SPHERE standards."* (Luqman, Zulqarnain and Bahadar, 2012, p.25).

Before the first final edition of the Sphere handbook was published in 2000, there was an intense debate on whether or not minimum technical standards in core areas of humanitarian response such as WASH shift the focus away from humanitarian principles and their underlying more difficult and pressing political issues (Buchanan-Smith, 2003, p.15). Back then, substantive research work[7] has been conducted to identify accountability and rights-based components to be included in the draft as a balance to the more quality-oriented

[7] Such research work includes e.g. *Towards Shared Social Learning for Humanitarian Programmes* (Apthorpe and Atkinson, 1999). This paper argues in favour of applying shared social learning to improve humanitarian interventions in terms of their appropriateness, effectiveness and accountability. Shared social learning is conceived as *"processes of working with people to discover and understand their own social and cultural ideas, realities and actual practices of meeting immediate survival needs in times of severe stress, such as coping mechanisms adopted including through relief provision"* (Apthorpe and Atkinson, 1999, p.5). Some of its intentions are now embodied in the Sphere core standards.

technical standards. The achieved compromise was the distinct rights-based structure of the Sphere standards explained in *Chapter 2*. And the Sphere standards, like humanitarian standards in general, are an on-going process of adaption and refinement. Probably because Sphere's philosophy is not fully grasped by the humanitarian community and the minimum technical standards are what Sphere is best known for, the latest edition of the Sphere handbook (2011) has put a standard on WASH programme design and implementation in the first place in the minimum standards in WASH.

This standard (*Figure 12*) explicitly states that disaster-affected people should be involved in the design, management and maintenance of the facilities where appropriate. Several of the previously presented key findings from the evaluation reports reveal that in many interventions it would have been appropriate but for unknown reasons people have not been involved. Evaluations could do a much better job in shedding light on the reasons why certain standards were or were not achieved. The additional information provided with the standard definitions offer a range of pointers where to start looking for and what should be considered when elaborating findings. It is through such kind of evidence that EHA can keep its stake in improving the performance of humanitarian action.

Because of above described misinterpretation of the Sphere minimum standards in WASH, only few relevant and valuable findings can be retrieved from the evaluation reports – this although, in total, nearly all minimum standards are referenced in the researched evaluations (see *Table 11* on the next page). In many instances it is hard to accomplish certainty whether or not a certain finding is conceived with a specific standard in mind. While this is a general methodological challenge particularly relevant for the analysis in this chapter, what matters in the end for EHA is that core elements of standards are used. With respect to the intentions behind the minimum standard definitions and the key findings in *Table 11* it can be summarised that,

- firstly, it is the responsibility of humanitarian actors to translate the Sphere minimum standards in WASH into concrete figures for monitoring coverage and effectiveness of interventions,
- secondly, if there is agreement on specific indicators to be used in the WASH sector (e.g. in the form of WASH cluster guidelines), it is of little value for learning purposes (in addition for decision makers) to use the whole evaluation effort when considering standards to pinpoint how difficult it is to achieve quantitative benchmarks or how inappropriate[8] they might be,
- and thirdly, also because monitoring with qualitative measures, i.e. standards according to their intention, is typically neglected during emergency response, a focus on them to establish evidence that matters for improving performance is suggested for EHA.

[8] What could be a realistic alternative? Simply lowering quantitative coverage measures (through for example agreeing that 100 persons per community latrine is acceptable considering the context instead of 20) so that humanitarian agencies can report they have achieved "standards" gets humanitarian emergency response nowhere.

Sphere Minimum Standards in WASH	
Standards/ references to standards used:	Access and water quantity; Appropriate and adequate toilet facilities (Grünewald, et al., 2011a, p.33) WASH programme design and implementation (Barham, Nabunny and Philpott, 2011a, p.18); Access and water quantity; Water quality (DeVillez, Bousquet and Nyagwambo, 2011, p.21); WASH programme design and implementation; Hygiene Promotion implementation; Access and water quantity; Environment free from human faeces; Appropriate and adequate toilet facilities; Individual and family protection (vector control); Physical, environmental and chemical protection measures (vector control); Collection and disposal of solid waste; Drainage work (Grayel and Mattson, 2012, pp.19-32); WASH programme design and implementation; Access and water quantity (Fox, 2008, pp.11-14); WASH programme design and implementation; Hygiene Promotion implementation; Environment free from human faeces; Appropriate and adequate toilet facilities; Individual and family protection (vector control); Collection and disposal of solid waste; Drainage work (Fortune and Rasal, 2010, pp.14-22); WASH programme design and implementation; Hygiene Promotion implementation; Access and water quantity; Water quality; Water facilities; Environment free from human faeces; Appropriate and adequate toilet facilities; Collection and disposal of solid waste (Luqman, Zulqarnain and Bahadar, 2012, pp.25-26); WASH programme design and implementation; Hygiene Promotion implementation; Access and water quantity; Water quality; (Laurens, 2005, pp.17-74); WASH programme design and implementation; Access and water quantity; Appropriate and adequate toilet facilities; Drainage work (Van der Wijk, Mwezi and Kazan, 2010, pp.26-28); Access and water quantity (Jeene, 2010, pp.7-18)
Selected key findings related to standards:	Common standards are inappropriate in urban contexts (Grünewald, et al., 2011a, p.33); The shift from Sphere standards to the flexible government standards where people are expected to maximally participate in the process of accessing water is not “well-thought-through” and will affect effectiveness (Barham, Nabunny and Philpott, 2011a, p.18); Lack of monitoring and agreement on the details of standards make an objective assessment of how efficiently targets were met impossible (DeVillez, Bousquet and Nyagwambo, 2011, p.21); “Step-by-step and participation-based research of local solutions […] while certainly would not have fixed all the complex issues it could have produced more [effective sanitation] alternatives.” (Grayel and Mattson, 2012, pp.19-32); “The reluctance of the population to pay for the water would change over time as the quality of the water is increasingly appreciated […] the information campaign on the value of the water was too short and inadequate” (Fox, 2008, p.14); “reports highlighted the ongoing risk of open defecation or defecation in plastic bags (such as by women at night) which were thrown onto the normal waste piles. The official message to people was not to dispose of these plastic bags with other solid waste. However, there did not seem to be any alternative for people, who were also not supposed to put the bags into the latrines. The lack of a solution for this issue appears to have been a gap in the […] response to reduce the risk of contamination from faeces.” (Fortune and Rasal, 2010, p.21); Sphere standards may not be relevant (Luqman, Zulqarnain and Bahadar, 2012, p.25); High level of community involvement should not be preferred over good workmanship (Van der Wijk, Mwezi and Kazan, 2010, p.27); “This intervention once again draws attention to the fact that host populations often have a water situation that is well below SPHERE standards, even before a crisis happens.” (Jeene, 2005, p.3)

Table 11: Sphere minimum standards in WASH: Summary of use and key findings

As the analysis in *Chapter 3* and above key findings (*Table 11*) show, the Sphere minimum standards in WASH cut across several OECD-DAC criteria. Some of the closest relations exist for the criteria relevance/appropriateness, coverage and effectiveness. But because they are confused with quantitative indicators or typically used as such, the evaluation reports only offer cursory insights into the relevant factors for achieving them. The minimum standards define core outputs of WASH emergency response and provide the "best-known indicators of humanitarian impact" in this sector (ECB, 2007, p.45). As such, they also "create a 'common language' and enable comparison between projects" (ibid.). The latter is a further reason why the Sphere minimum standards in WASH (where applicable according to the intervention) should be considered at the heart of creating evidence on what works best in which context[9].

As explored in *Chapter 2*, qualitative and quantitative methods are required in evaluations to create evidence with the latter being identified as less well used. Keeping in mind the overall goal of WASH disease prevention (*Figure 2*) in emergency interventions, quantifiable indicators actually play a dominant role in proving health impacts. In this regard, the scope of the Sphere minimum standards in WASH is of particular concern. As a recent evidence review of research on health interventions in humanitarian crises (Blanchet, et al., 2013) suggests, providing credible evidence on health impacts through WASH interventions is currently beyond the scope of scientific as well as evaluative research. The research needs identified by the study are vast (Blanchet, et al., 2013, pp.43-47). With respect to the Sphere minimum standards in WASH Blanchet, et al. (2013, p.47) state: *"Sphere indicators are important but difficult to measure in practice. A review of Sphere indicators in the WASH sector is needed. Without good indicators it is impossible to know how well the interventions have been working in relation to acceptable levels (e.g. diarrhoea) given a setting, population, and specific disease (e.g. Shigella vs. cholera)."*.

So, even if the Sphere minimum standards in WASH are used according to their intentions, it would be hardly possible on their basis to establish evidence on how effective an intervention has been in preventing WASH related diseases. For current EHA this suggest to focus on the kind of evidence that is possible to achieve. Up to this point, the analysis has drawn attention to several WASH gaps and learning needs that are within the scope of EHA considering its possibilities as well as constraints but remain less well addressed. Although the Sphere and HAP standards are not perfect evaluation criteria, they can add valuable and needed quality and accountability aspects to the OECD-DAC criteria. Finally, being effective in contributing to WASH disease prevention is also about better engagement with disaster-affected people. As noted introductory and evident from the meta-analysis, experiences made often receive scant attention in greater detail, leaving little to inform future

---

[9] The earlier noted *Model Guidelines for Mainstreaming Water and Sanitation in Emergencies, Protracted Crises, LRRD and Disaster Preparedness Operations* (DG ECHO, 2005a) include a generic guide on the flexible application of the Sphere core standards and minimum standards in WASH. It presents the extent to which the standards' key indicators are likely to be adjusted according to three principal scenarios: Acute emergency responses, chronic or post-acute emergency responses and displaced populations in camps. Although based on the 2004 version of the standards, the guide offers a valuable resource for evaluators regarding the choice and adaption of Sphere indicators in different emergency contexts.

response. The *2013 Humanitarian Accountability Report* concludes: *"What is remarkable is that after 10 years, the questions and criticisms raised by disaster-affected communities continue to be the same time and time again."* (Darcy, Alexander and Kiani, 2013, p.71). However, as pointed out in *Chapter 2*, EHA can only be part of any solution.

5 Summary and Conclusions

5.1 Summary of Research Findings

The thesis examined how the OECD-DAC framework is applied in evaluations of humanitarian response to disasters in the WASH sector. At the centre of the research has been a meta-analysis of evaluation reports reflecting the relief to development contiguum. The guiding research questions address the main components of this framework as suggested by EHA guides – the OECD-DAC evaluation criteria, cross-cutting themes and humanitarian standards. In the focus have been how the OECD-DAC criteria as its core elements are used in conjunction with recommended cross-cutting themes and relevant standards, what they reveal, and finally, which benefits and shortcomings can be identified regarding their use for evaluating WASH emergency interventions. As humanitarian emergencies are complex, unique and dynamic, EHA can play an important role in deriving lessons from field experience to establish evidence on *What works, what not and why?* in different contexts. Considering that informing future response is the primary motivation for EHA, the lenses through which humanitarian work is conceived matter, albeit of course how evaluation criteria are used defines the quality of evaluations rather than what criteria are employed.

As explored in *Chapter 2*, substantive WASH gaps have been identified by the GWC and evaluation is considered as a tool to address learning needs, but the basic conditions for EHA remain suboptimal to be used as such. This is evident through literature as well as the researched evaluation reports. Besides that knowledge management and quality assurance methods following up evaluation to assure learning and performance improvements remain weak or inexistent, generating evidence through EHA is confronted with several general constraints. Evaluations typically take place in data-poor and operationally fluid environments. While the latter will always pose certain difficulties hard to overcome, humanitarian agencies could contribute much more to minimise potential challenges for EHA – through e.g. making available the resources needed for response monitoring or proper and timely planning of evaluations. In addition, evaluations (except for RTEs) are expected to extract relevant findings from a complex of issues after long intervention periods (often more than one year) and the evaluation purpose (accountability or learning) remains rather unclear, like the expected use of the OECD-DAC evaluation framework.

The OECD-DAC evaluation criteria are dominant in EHA, but their use is not always well understood, having led to poor evaluation results. In addition to the OECD-DAC criteria's lack of covering the quality of process, researchers (e.g. De Geoffroy and Kauffmann, 2009) have pointed out the need for having easier to handle evaluation criteria, also considering that a quality framework for evaluation should be a set of common values shared between all stakeholders; not least to promote a more active role of disaster-affected people in EHA. In humanitarian work, such common values already exist in the form of standards which organisations have committed them to on a voluntary basis. For WASH, Sphere and HAP can be identified as the two most important standard initiatives, with the former providing the only WASH specific standards so far established for emergency response.

The Sphere and HAP standards are elaborated and complementary frameworks, together addressing quality and accountability in a comprehensive manner. Following recent EHA guides (e.g. Buchanan-Smith and Cosgrave, 2013) standards themselves can be suggested as evaluation criteria, and as an alternative to the OECD-DAC criteria. In the researched evaluations, the Sphere and Hap standards are used with varying extent and intent together with the OECD-DAC criteria. As the key findings from these evaluations (*Chapter 3 and 4*) demonstrate, aspects of both are relevant, in particular to address the two fundamental and interrelated weaknesses of WASH emergency response: LRRD and involvement of disaster-affected people and communities at all stages, from assessment to the implementation of interventions.

Considering this as well as the general need to make evaluation frameworks less complex and confusing, the idea of having a limited set of generic evaluation criteria as convenient entry points into evaluation, and perhaps in parallel a number of “sub-criteria” adapted to various situations – as suggested by ECHO (DG ECHO, 2007, p.51) – is worth being explored for the evaluation of WASH emergency response. To be able to draw further conclusions for this approach, in the following, the main findings from the meta-analysis are summarised – thereby answering the first two and closely related research questions: *How are the OECD-DAC criteria used to evaluate humanitarian emergency response in the WASH sector and what do they reveal? How are cross-cutting themes and standards used in conjunction with the OECD-DAC criteria?*

The evaluation reports reflect a variety of uses of the OECD-DAC criteria ranging from focussing on a few criteria under which aspects of others are subsumed, over addressing selected aspects of criteria where relevant under the heading of evaluation questions, to a rather mechanistic application of the whole set. Based on what the criteria are used to reveal, a few core evaluation criteria can be identified under which aspects of other criteria are subsumed in the researched evaluations: *connectedness, efficiency and effectiveness*. *Connectedness* is a key criterion with which the complexity of LRRD is addressed and it allows to view findings related to other criteria together with standards and cross-cutting issues such as DRR and environment. Similarly, *efficiency* is a highly relevant criterion, especially considering the limited availability of resources (financial, human, technical and material). It serves more than a sometimes purely economic perceived value for money analysis, which is anyhow likely to be hampered by lack of data. Finally, *effectiveness* is probably the most important OECD-DAC criterion next to connectedness. A central quality of effectiveness is the weighing of different aspects against each other (e.g. cost-effective realisation vs. prototype interventions to ensure learning during implementation before going to scale).

Although *impact* would in theory be the most important criterion, in practice, the impact of concern, i.e. health impact of WASH interventions, is nearly impossible to assess. All other potential impacts of a response are better viewed with connectedness or can be encompassed with cross-cutting issues as well as the Sphere and HAP standards. For the purposes *coherence* is used in evaluations, it can be said that connectedness or the Sphere and HAP standards would be much better suited to evaluate these. It still gets confused with coordination by some evaluators and given is intended use, it can be proposed as a "sub-criteria" for politicised emergency contexts and sector-wide joint evaluations. Finally, *coverage and relevance/appropriateness* are criteria with strong causal links to effectiveness (appropriateness is also linked to connectedness in several evaluations), hence their aspects are (and in general can be) also captured with it. Additionally, the findings conceived under relevance/appropriateness reveal at large what the Sphere and HAP standards express (e.g. the importance of involving disaster-affected communities in the response). And coverage is almost exclusively based on the Sphere minimum standards in WASH and related indicators.

Overall, the key findings from the evaluation reports are mostly consistent with general findings from meta-evaluations covering the whole humanitarian sector (Stokke, 2007; Harvey, et al., 2010; Taylor, et al., 2013) and add little value on top of what the Sphere and HAP standards emphasise. While some of the identified global WASH gaps are addressed in evaluations, related findings remain below what can be considered an evidence base for learning. Especially regarding LRRD issues, the elaboration of findings gets rarely beyond general statements drawing attention to e.g. the importance of realistic and timely planned exit strategies.

The integration of cross-cutting themes is a common feature in all evaluations. Based on their use, four different thematic areas relevant to WASH can be distinguished: (i) gender and protection; (ii) environment; (iii) resilience; DRR and capacity building; (iv) accountability, participation, the inclusion of marginalised/vulnerable groups and similar themes which express key contents of the Sphere and HAP standards. Cross-cutting themes are typically used to voice major concerns, e.g. the need to source wood internationally for large scale latrine construction programmes in countries with already high levels of deforestation to not contribute to environmental degradation and future disasters. While adding important quality aspects to the scope of the OECD-DAC criteria, cross-cutting themes are just generic expressions of issues to be considered in humanitarian response.

Because the Sphere and HAP standards already include above themes at large, they can be suggested as a practical tool to maintain a focus on the quality and accountability aspects in EHA. However, similar to cross-cutting themes, the Sphere and HAP standards remain cursorily and inconsistently used in evaluations. In addition, qualitative standards are regularly mistaken as quantitative indicators and with few exceptions not used according to their intentions. Despite the Sphere and HAP standards address issues relevant to connectedness, efficiency and effectiveness, their integration is weak and findings conceived under these criteria rather reflect what standards convey. To sum up, the Sphere and HAP standards provide a shared quality and accountability framework that targets key weaknesses of humanitarian response. The comprehensive guidance developed for their use offers a valuable resource for evaluators to elaborate more informative findings, even though its suggested indicators are not always straightforward to operationalise.

5.2 Conclusions

This research work targeted a very specific issue out of a complex of problems EHA is confronted with. Against this background, the choice and use of evaluation criteria is not the most pressing issue for evaluating humanitarian response. With one of the latest EHA guides (Buchanan-Smith and Cosgrave, 2013, p.53) it can be summarised that criteria are just “tools to think with”, and finally what matters are the questions an evaluation wants to answer, not the criteria. But in general there is always the need to have an underlying evaluation framework specifying a set of criteria with guidance on their use to make sure that important aspects affecting the performance of humanitarian action are not left out. The OECD-DAC framework – as the common framework for EHA in the WASH sector – requires to be refocused to fulfil that purpose. Its main elements, the OECD-DAC criteria, have certain benefits as well as shortcomings. From the conducted analysis, the following conclusions can be drawn, thereby answering the third and main research question reflected in the title of the thesis:

- The nature of the OECD-DAC criteria is complex. They require a good understanding and experience in their use to be applied in a meaningful way. Integrating cross-cutting themes and particularly humanitarian standards to enhance their quality dimensions is not straightforward.
- The use of the OECD-DAC criteria largely reveals what is already well known and framed with the Sphere and HAP standards. Hence, these standards should be a more serious concern in evaluations to dig a level deeper and find out about the causes why certain standards which express core objectives of emergency response have not been achieved or adhered to.
- Despite these shortcomings, the OECD-DAC criteria still contain highly relevant and to some extent irreplaceable angles for EHA. Connectedness, efficiency and effectiveness have been identified as key evaluation criteria for WASH interventions. Also the use of this criteria is well understood.

In continuing where the summary of research findings has ended, the proposed generic evaluation criteria for WASH emergency response are: *Connectedness, efficiency, effectiveness and the Sphere and HAP standards*. This approach is slightly different from the one suggested by ECHO (*Chapter 2*) but seems more appropriate in light of persistent weaknesses and knowledge gaps in the WASH sector. The rationale behind it can be conveyed with the old maxim: *If you change the way you look at things, the things you look at change*. However, while on the one hand it can be argued that a focus on a few OECD-DAC criteria would allow to give way to a more comprehensive and consistent consideration of standards in evaluations, on the other hand, the concept of particularly the Sphere standards is not always understood well and according to its intentions within the EHA community. Training of not just humanitarian workers but also evaluators in standards seem necessary.

The research findings further point out the need to (re)focus evaluations on the whole if they are expected to deliver findings relevant for informing future response. Putting too much thematic scope in an evaluation will not serve any purpose EHA may have. Most notably when evaluations are supposed to provide an evidence base for learning, commissioning agencies should be very realistic about what an evaluation can contribute considering the constraints it is likely to face. The thesis has argued strongly in favour of using evaluations with this intention as evaluations focusing on WASH are rare but learning needs are vast. However, as the analysis has demonstrated, creating credible evidence is already limited at the level of effectiveness because of lack of data. Certainly, more and more focused evaluations will be needed to address the identified WASH gaps, but what is also required is a much closer relation between the functions of evaluation and emergency response. The emerging practice of conducting evaluations while interventions are implemented is already pointing in this direction.

Finally, to address the probably most fundamental weakness of WASH emergency interventions, linking relief and development, a much needed perspective is the one of disaster-affected communities. As of now, involving primary stakeholders in setting the evaluation agenda would be a novelty, but demonstrate a much closer understanding of what accountability involves. To achieve a more active role of disaster-affected people in EHA, the Sphere and HAP standards can provide a valuable roadmap. In fact, many of their indicators are ideally assessed by affected people than external evaluators.

Still, like the OECD-DAC criteria or any other evaluation criteria, the Sphere and HAP standards are not perfect. This is particularly evident with respect to the main aim of WASH interventions: WASH disease prevention through hygiene improvement in emergencies. As noted in the analysis, a review of the indicators of the Sphere minimum standards in WASH is needed to be able to create evidence on how effective interventions are in preventing specific diseases. How to establish evidence on health impacts through WASH interventions is definitely the most pressing area for future research. Recommendations for future research, specifically addressing EHA, are not provided here but in detail by Blanchet, et al. (2013, p.47). Considering the impact of the Sphere and HAP standards so far, getting closer to evaluating impact could only bring about much-needed change in how WASH emergency response is conducted if the incentives for humanitarian actors to change their practices are existent.

6 References

ACF IN (Action Against Hunger International Network), 2008. *How to make WASH projects sustainable and successfully disengage in vulnerable contexts*. [pdf] Available at: <http://www.actionagainsthunger.org/sites/default/files/publications/Sustainability_of_WASH_projects_-_Guidelines.pdf> [Accessed 15 October 2013].

ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action), 2003. *Annual Review 2003. Humanitarian Action: Improving Monitoring To Enhance Accountability and Learning*. London: Overseas Development Institute.

ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action), 2005. *Assessing the Quality of Humanitarian Evaluations, The ALNAP Quality Proforma 2005*. [pdf] Available at: <http://www.alnap.org/pdfs/QualityProforma05.pdf> [Accessed 27 September 2013].

Apthorpe, R. and Atkinson, P., 1999. *Towards Shared Social Learning for Humanitarian Programmes*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pdfs/other_studies/rashared.pdf> [Accessed 27 September 2013].

Bastable, A. and Russell, L., 2013. *Gap Analysis in Emergency Water, Sanitation and Hygiene Promotion*. [pdf] s.l.: Humanitarian Innovation Fund. Available at: <http://www.washcluster.info/?q=download/file/fid/637> [Accessed 20 October 2013].

Beck, T., 2006. *Evaluating humanitarian action using the OECD-DAC criteria. An ALNAP guide for humanitarian agencies*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pool/files/eha-2006.pdf> [Accessed 27 September 2013].

Beck, T. and Borton, J., eds., 2004. *ALNAP Review of Humanitarian Action in 2003: Field Level Learning*. London: Overseas Development Institute.

Blanchet, K., Sistenich, V., Ramesh, A., Frison, S., Warren, E., Hossain, M., Knight, A., Lewis, C., Smith, J., Woodward, A., Dahab, M., Pantuliano, S. and Roberts, B., 2013. *An evidence review of research on health interventions in humanitarian crises. Final Report*. [pdf] Available at: <http://www.elrha.org/uploads/Evidence%20Review%20R2HC%20Full%20Report%20FINAL_2.pdf> [Accessed 27 November 2013].

Borton, J. and Robertson, K., eds., 2002. *ALNAP Annual Review 2002. Humanitarian Action: Improving Performance Through Improved Learning*. London: Overseas Development Institute.

Buchanan-Smith, M., 2003. *How the Sphere Project Came into Being: A Case Study of Policy-Making in the Humanitarian Aid Sector and the Relative Influence of Research*. [pdf] London: Overseas Development Institute. Available at: <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/176.pdf> [Accessed 27 September 2013].

Buchanan-Smith, M. and Cosgrave, J., 2010? .*Evaluation of Humanitarian Action (EHA) – Course Reference Manual. Version 1.30*. [pdf] Available at: <http://www.alnap.org/pool/files/2010-training.pdf> [Accessed 27 September 2013].

Buchanan-Smith, M. and Cosgrave, J., 2013. *Evaluation of Humanitarian Action (EHA). Pilot Guide*. [pdf] Available through: The Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) website <http://www.alnap.org/> [Accessed 27 September 2013].

Chianca, T., 2008. The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. *Journal of MultiDisciplinary Evaluation*, 5(9), pp.41-51.

Collinson, S. and Duffield, M., 2013. *Paradoxes of presence. Risk management and aid culture in challenging environments*. [pdf] London: Overseas Development Institute. Available at: <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8428.pdf> [Accessed 27 September 2013].

Cosgrave, J., Ramalingam, B. and Beck, T., 2009. *Real-time evaluations of humanitarian action. An ALNAP Guide*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pool/files/rteguide.pdf> [Accessed 27 September 2013].

Darcy, J., Alexander, J. and Kiani, M., 2013, *2013 Humanitarian Accountability Report*. [pdf] London: Humanitarian Accountability Partnership. Available at: <http://www.hapinternational.org/pool/files/2013-har.pdf> [Accessed 20 October 2013].

Darcy, J. and Knox Clarke, P., 2013. *Background paper – 28th ALNAP Meeting on Evidence & Knowledge in Humanitarian Action*. Washington, D.C., United States of America, 5-7 March 2013. [pdf] Available at: <http://www.alnap.org/pool/files/background-paper-28th-meeting.pdf> [Accessed 27 September 2013].

Darcy, J., Stobaugh, H., Walker, P. and Maxwell, D., 2013. *The Use of Evidence in Humanitarian Decision Making*. [pdf] Somerville, MA: Feinstein International Center. Available at: <http://sites.tufts.edu/feinstein/files/2013/02/TUFTS_1306_ACAPS_3_online.pdf> [Accessed 27 September 2013].

De Geoffroy, V. and Kauffmann, D., 2009. Building a quality framework: issues and challenges. *Humanitarian Aid on the Move*, [online] Available at: <http://www.urd.org/newsletter/spip.php?article52> [Accessed 15 October 2013].

DG ECHO (European Commission Directorate-General for Humanitarian Aid), 2005a. *Model Guidelines for Mainstreaming Water and Sanitation in Emergencies, Protracted Crises, LRRD and Disaster Preparedness Operations*. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/watsan2005/files/model_guidelines.pdf> [Accessed 15 October 2013].

DG ECHO (European Commission Directorate-General for Humanitarian Aid), 2005b. *Concept Paper for Mainstreaming Water and Sanitation in Emergencies, Protracted Crises, LRRD and Disaster Preparedness Operations*. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/watsan2005/files/concept_paper.pdf> [Accessed 15 October 2013].

DG ECHO (European Commission Directorate-General for Humanitarian Aid), 2007. *Evaluation of Humanitarian Aid by and for NGOs. A guide with ideas to consider when designing your own evaluation activities*. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/2007/humanitarian_guide.pdf> [Accessed 15 October 2013].

ECB (The Emergency Capacity Building Project), 2007. *Impact Measurement and Accountability in Emergencies: The Good Enough Guide*. [pdf] Oxford: Oxfam GB. Available at: <http://www.ecbproject.org/downloads/resources/good-enough-guide-book-en.pdf> [Accessed 27 September 2013].

EHP (Environmental Health Project), 2004. *The Hygiene Improvement Framework. A Comprehensive Approach for Preventing Childhood Diarrhea.* [pdf] Washington, D.C.: Environmental Health Project. Available at: <http://www.ehproject.org/PDF/Joint_Publications/JP008-HIF.pdf> [Accessed 10 December 2013].

Eggers, H.W., 2009. Comments and Proposals Concerning Chianca's "The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement". *Journal of MultiDisciplinary Evaluation*, 6(11), pp.116-124.

Few, R., McAvoy, D., Tarazona, M. and Walden, V.M., 2014. *Contribution to Change: An Approach to Evaluating the Role of Intervention in Disaster Recovery*. [pdf] Rugby, UK: Practical Action Publishing and Oxford: Oxfam GB. Available at: <http://www.ecbproject.org/downloads/contribution-to-changeweb-pdf-updated-11-12-2013.pdf> [Accessed 10 December 2013].

Good Humanitarian Donorship (GHD) Initiative, 2003. *Principles and Good Practice of Humanitarian Donorship*. Available through: The Good Humanitarian Donorship website at: <http://www.goodhumanitariandonorship.org> [Accessed 20 November 2013].

Grünewald, F., Kauffmann, D., Boyer, B. and Patine, J., 2011b. *Humanitarian aid in urban settings: current practice, future challenges*. [pdf] Available at: <http://www.alnap.org/pool/files/rapport-echo-urba-alnap-vfinale-aout.pdf> [Accessed 27 September 2013].

Guerrero, S., Woodhead, S. and Hounjet, M., 2013. *On the right track? A brief review of monitoring and evaluation in the humanitarian sector*. [pdf] s.l.: Action Against Hunger/Consortium of British Humanitarian Agencies. Available at: <http://www.actionagainsthunger.org.uk/fileadmin/contribution/pdf/On%20the%20Right%20Track.pdf> [Accessed 27 September 2013].

GWC (Global WASH Cluster), 2009a. *WASH Accountability Checklist*. [pdf] Available at: <http://www.ecbproject.org/downloads/Accountability/WASH_Accountability_Checklist.pdf> [Accessed 27 September 2013].

GWC (Global WASH Cluster), 2009b. *WASH Accountability Resources: Ask, Listen, Communicate*. [pdf] New York: Global WASH Cluster. Available at: <https://www.humanitarianresponse.info/system/files/documents/files/wash-accountability-handbook.pdf> [Accessed 27 September 2013].

GWC (Global WASH Cluster), 2011. *Global WASH Cluster Strategic Plan 2011 – 2015*. [pdf] Available at: <http://www.washcluster.info/?q=download/file/fid/316> [Accessed 27 September 2013].

GWC (Global WASH Cluster), 2013. *17th Global WASH Cluster Meeting. Minutes of Meeting*. [pdf] Available at: <http://www.washcluster.info/?q=download/file/fid/645> [Accessed 27 September 2013].

Hallam, A., 1998. *Evaluating Humanitarian Assistance Programmes in Complex Emergencies*. [pdf] London: Overseas Development Institute. Available at: <http://www.odihpn.org/download/gpr7pdf> [Accessed 27 September 2013].

Hallam, A., 2011. *Harnessing the Power of Evaluation in Humanitarian Action: An initiative to improve understanding and use of evaluation*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pool/files/evaluation-alnap-working-paper.pdf> [Accessed 27 September 2013].

Hallam, A. and Bonino, F., 2013. *Using Evaluation for a Change: Insights from humanitarian practitioners. ALNAP Study*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pool/files/alnap-study-using-evaluation-for-a-change.pdf> [Accessed 20 October 2013].

Harries, K., 2013. *Knowledge Management in Practice. Implementing Effective Knowledge Management in Emergencies: A Case Study from Somalia WASH Cluster*. [pdf] Available at: <http://recaproject.org/wp-content/uploads/2013/02/Knowledge-Management-in-Action.pdf> [Accessed 15 October 2013].

Harvey, P., Stoddard, A., Harmer, A. and Taylor, G., 2010. *The State of the Humanitarian System: Assessing Performance and Progress- A Pilot study*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pool/files/alnap-sohs-final.pdf> [Accessed 27 September 2013].

Humanitarian Accountability Partnership (HAP) International, 2013. *Guide to the 2010 HAP Standard in Accountability and Quality Management*. [pdf] Geneva: HAP International. Available at: <http://www.hapinternational.org/pool/files/guide-to-the-2010-hap-standard-printer-friendly-version-new.pdf> [Accessed 20 October 2013].

IFRC (The International Federation of Red Cross and Red Crescent), 2011a. *IFRC Framework for Evaluation*. [pdf] Available at: <http://www.ifrc.org/Global/Publications/monitoring/IFRC-Framework-for-Evaluation.pdf> [Accessed 27 September 2013].

IFRC (The International Federation of Red Cross and Red Crescent), 2011b. *Project/programme monitoring and evaluation (M&E) guide*. [pdf] Available at: <http://www.ifrc.org/Global/Publications/monitoring/IFRC-ME-Guide-8-2011.pdf> [Accessed 27 September 2013].

Lindahl, C., 1998. *The Management of Disaster Relief Evaluations. Lessons from a Sida evaluation of the complex emergency in Cambodia*. [pdf] Stockholm: Swedish International Development Cooperation Agency. Available at: <http://www.alnap.org/pool/files/46228270.pdf> [Accessed 27 September 2013].

Maury, H., 2009. What is evaluation for? The missing links in the evolution of humanitarian actors. *Humanitarian Aid on the Move*, [online] Available at: <http://www.urd.org/newsletter/spip.php?article59> [Accessed 15 October 2013].

Proudlock, K. and Ramalingam, B., 2009. Improving humanitarian impact assessment: bridging theory and practice. In: ALNAP 2009. *ALNAP 8th Review of Humanitarian Action: Performance, Impact and Innovation*. London: Overseas Development Institute. [pdf] Available at: <http://www.alnap.org/pool/files/8rhach2.pdf> [Accessed 27 September 2013].

Ramalingam, B. and Mitchell, J., 2009. Counting what counts: performance and effectiveness in the humanitarian sector. In: ALNAP 2009. *ALNAP 8th Review of Humanitarian Action: Performance, Impact and Innovation*. London: Overseas Development Institute. [pdf] Available at: <http://www.alnap.org/pool/files/8rhach1.pdf> [Accessed 27 September 2013].

Ramalingam, B., Scriven, K. and Foley, C., 2009. Innovations in international humanitarian action. In: ALNAP 2009. *ALNAP 8th Review of Humanitarian Action: Performance, Impact and Innovation*. London: Overseas Development Institute. [pdf] Available at: <http://www.alnap.org/pool/files/alnap-rha-2009.pdf> [Accessed 27 September 2013].

Raynard, P., 2000. *Mapping Accountability in Humanitarian Assistance*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pdfs/other_studies/praccountability.pdf> [Accessed 27 September 2013].

Rogers, P.J., 2012. *Introduction to impact evaluation*. [pdf] s.l.: InterAction. Available at: <http://www.interaction.org/sites/default/files/1%20-%20Introduction%20to%20Impact%20Evaluation.pdf> [Accessed 27 September 2013].

Sandison, P., 2006. The utilisation of evaluations. In: J. Mitchell ed. 2006. *ALNAP Review of Humanitarian Action: Evaluation utilisation*. London: Overseas Development Institute. pp.89-144. [pdf] Available at: <www.alnap.org/pool/files/ch3-f1.pdf> [Accessed 27 September 2013].

Stokke, K., 2007. *Humanitarian Response to Natural Disasters: A Synthesis of Evaluation Findings*. [pdf] Oslo: Norwegian Agency for Development Cooperation. Available at: <http://www.norad.no/en/tools-and-publications/publications/publication/_attachment/107610?_download=true&_ts=11ebb1a4196> [Accessed 27 September 2013].

Taylor, G., Stoddard, A., Harmer, A., Haver, K. and Harvey, P., 2012. *The State of the Humanitarian System 2012 Edition*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pool/files/alnap-sohs-2012-lo-res.pdf> [Accessed 27 September 2013].

The Sphere Project, 2009. *Taking the initiative. Exploring quality and accountability in the humanitarian sector: an introduction to eight initiatives*. [pdf] Geneva: The Sphere Project. Available at: <http://www.sphereproject.org/download/51828bace1d9f> [Accessed 20 October 2013].

The Sphere Project, 2011. *Humanitarian Charter and Minimum Standards in Humanitarian Response*. 3rd ed. Bourton on Dunsmore, UK: Practical Action Publishing.

Turnbull, M., Sterrett, C. L. and Hilleboe, A., 2013. *Toward Resilience. A Guide to Disaster Risk Reduction and Climate Change Adaption*. [pdf] Rugby, UK: Practical Action Publishing. Available at: <http://www.ecbproject.org/downloads/ECB-toward-resilience-Disaster-risk-reduction-Climate-Change-Adaptation-guide-english.pdf> [Accessed 20 October 2013].

Van Brabant, K., 1997. *Organisational and Institutional Learning in the Humanitarian Sector – Opening the Dialogue*. [pdf] London: Overseas Development Institute. Available at: <http://www.alnap.org/pdfs/other_studies/kvblearn.pdf> [Accessed 27 September 2013].

Evaluation Reports

Barham, B., Nabunnya, H. and Philpott, J., 2011a. *Evaluation of DG ECHO'S Action in Uganda, Evaluation Report*. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/2011/Uganda_Evaluation_report.pdf> [Accessed 27 September 2013].

Barham, B., Nabunnya, H. and Philpott, J., 2011b. *Evaluation of DG ECHO'S Action in Uganda, Linking Relief and Rehabilitation to Development Report*. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/2011/Uganda_LRRD_Final_report.pdf> [Accessed 27 September 2013].

DeVillez, P., Bousquet, C. and Nyagwambo, N.L., 2011. Evaluation of DG ECHO's Action in the Water and Sanitation/Public Health Sector in Zimbabwe, Main Report. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/2011/Zimbabwe_Final_Report.pdf> [Accessed 27 September 2013].

Fortune, V. and Rasal, P., 2010. *British Red Cross – Mass Sanitation Module: 2010 Haiti Earthquake Response. Post Deployment Learning Evaluation*. [pdf] Available at: <http://www.alnap.org/pool/files/haiti-msm-evaluation-report-final-020810.pdf> [Accessed 27 September 2013].

Fox, P., 2008. *Phase 2 Evaluation Report. Supplementary Water Supply System. IFRC, The Maldives*. [pdf] Available at: <http://www.ifrc.org/docs/Evaluations/eval08/AsiaPacific/MVSupplementaryWaterSupplyPhase2.08.pdf> [Accessed 27 September 2013].

Grayel, Y. and Mattson, K., 2012. *International Federation of Red Cross and Red Crescent Societies (IFRC) Water, Sanitation, and Hygiene Promotion: Final Evaluation Haiti Earthquake Operation*. [pdf] Available at: <http://adore.ifrc.org/Download.aspx?FileId=46644&.pdf> [Accessed 27 September 2013].

Grünewald, F., Kauffmann, D., Boyer, B. and Patine, J., 2011a. *Real-time evaluation of humanitarian action supported by DG ECHO in Haiti, 2009 – 2011*. [pdf] Available at: <http://ec.europa.eu/echo/files/evaluation/2011/Groupe-URD_evaluation_Haiti_en.pdf> [Accessed 27 September 2013].

Jeene, H., 2010. *CAFOD/CDTY Emergency Intervention in Western Equatoria, Southern Sudan, 2009-2010. Water and Health Promotion Component*. [pdf] Available at: <http://www.alnap.org/pool/files/evaluation-cafod-echo-2010-external-evaluation.pdf> [Accessed 27 September 2013].

Laurens, L., 2005. *Action Contre La Faim, Liberia. Dutch Water and Sanitation Programme in Bong County 2004-2005. External Evaluation of the Social Approach of the Water and Sanitation Programme. Final Report*. [pdf] Available at: <http://www.alnap.org/pool/files/liberia-final-report-external-evaluation-of-the-social-approach-watsan-bong-sep-23,-2005.pdf> [Accessed 27 September 2013].

Luqman, A., Zulqarnain, M. and Bahadar, N., 2012. *External Evaluation of ACF's Humanitarian Support to Conflict and Flood-affected Populations in Pakistan*. [pdf] Available at: <http://www.alnap.org/pool/files/acf-pakistan-kpha-programme-evaluation-report-annexes-2.pdf> [Accessed 27 September 2013].

Van der Wijk, J., Mwezi, E. B. and Kazan, L., 2010. *Evaluation of the DEC-funded CAFOD Health and WASH Project in the DRC*. [pdf] Available at: <http://www.alnap.org/pool/files/dec-drc-final-evaluation-report-23-march-2010.pdf> [Accessed 27 September 2013].

Websites

ALNAP	http://www.alnap.org/ ALNAP's official website with information and resources on accountability and performance of humanitarian action including an evaluation reports data base.
Groupe URD	http://www.urd.org Groupe Urgence, Réhabilitation and Développement (URD) provides resources on a broad range of evaluative and operational research of current themes in the emergency and development sector.
HumanitarianResponse.info	https://www.humanitarianresponse.info A service of the United Nations Office for the Coordination of Humanitarian Affairs (UN-OCHA), providing resources for humanitarian coordination, cluster specific information and access to country specific emergency sites.
Joint Standards Initiative (JSI)	http://www.jointstandards.org A process launched by the Humanitarian Accountability Partnership (HAP), People In Aid and the Sphere Project, in response to the perceived confusion, lack of awareness and inconsistent application of standards, and to seek greater coherence for users of standards.

Annexes

Annex 1 The ALNAP Quality Proforma

ASSESSING THE QUALITY OF HUMANITARIAN EVALUATIONS

THE ALNAP QUALITY PROFORMA 2005 (v. 02/03/05)

1. Background

ALNAP developed this Quality Proforma in 2000/2001 as a way of assessing humanitarian evaluation reports drawing on current thinking and good practice in the evaluation of humanitarian action.[1]
The overall aim of the Quality Proforma is to improve the quality of humanitarian evaluation practice. It does this by:

1. Providing an assessment tool for ALNAP's annual meta-evaluation of humanitarian evaluation reports as part of its *Review of Humanitarian Action*[2] series. The meta-evaluation seeks to identify trends in the quality of humanitarian evaluations, identifying both good and weak practices.[3]

2. Providing a checklist for evaluation managers and evaluators.

The Quality Proforma has undergone refinements during its application in four ALNAP *Reviews* between 2001 and 2003/4, in order to strengthen consistency in interpretation and usage and reflect developments in current thinking in the evaluation of humanitarian action. This version of the Proforma has undergone a process of simplification and reordering for the *Review of Humanitarian Action in 2004* in order to make it more accessible.

2. Meta-evaluation process

Each evaluation report included in ALNAP's meta-evaluation is rated against the Quality Proforma by two assessors working independently. For each report, every area of the criteria is given a comment and a rating. The ratings are then used to assess strengths and weaknesses of the set as a whole.
Since 2003/4, the draft findings of the Quality Proforma assessments have been discussed with a selection of the commissioning agencies in order to better understand the background to the evaluation process, gather information that may not show up in the written report and stimulate agency involvement and interest. The outcome of these discussions may lead to revisions of the final assessments. In 2005 for the first time, a selection of evaluators will also be consulted on the evaluation processes.

3. Using the ALNAP Quality Proforma

The development of the Proforma is linked to ALNAP's definition of the Evaluation of Humanitarian Action (EHA) given in the box below.
The Proforma is intended to be used for reports dealing with natural disasters and complex political emergencies. It should also be of value for other types of evaluative exercises in the humanitarian context. Although originally designed with programme evaluations in mind, the Proforma can also be used to review evaluations of such activities as humanitarian management processes, funding partnerships and sectoral approaches. In these cases, some questions in the Proforma may be noted as not relevant.

ALNAP Definition of Evaluation of Humanitarian Action (EHA)
"A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice, and enhance accountability. It has the following characteristics: i). it is commissioned by or in cooperation with the organisation(s) whose performance is being evaluated; ii). it is undertaken either by a team of non-employees (external) or by a mixed team of non-employees (external) and employees (internal) from the commissioning organisation and/or the organisation being evaluated; iii). it assesses policy and/or practice against recognised criteria (e.g., efficiency, effectiveness/timeliness/coordination, impact, connectedness, relevance/appropriateness, coverage, coherence and as appropriate, protection); and, iv). it articulates findings, draws conclusions and makes recommendations." ALNAP 2001, *Humanitarian Action: Learning from evaluation,* ALNAP Annual Review 2001. London: ALNAP/ODI.

The Quality Proforma is divided into six sections:

1. Assessing the Terms of Reference;
2. Assessing Evaluation Methods, Practice and Constraints;
3. Assessing Contextual Analysis;
4. Assessing the Intervention;
5. Assessing the Report;
6. Overall Comments.

Each section has four column headings:

- **Area of Enquiry** (subject matter)
- **Guidance Notes** (guidance as to what is deemed 'satisfactory' to ensure a degree of consistency of interpretation)
- **Comments** (to include a brief reason for the rating given)
- The **Rating**.

The rating system used for the meta-evaluation is as follows: A = Good
B = Satisfactory
C = Unsatisfactory
D = Poor
Z = Not applicable. (Where an area of enquiry is deemed not applicable, reasons should be given in the 'Comments' column. The proforma user's judgement remains a central factor in the rating exercise.)

Where the Guidance Note lists a number of areas that should be covered for an Area of Enquiry, a 'B' (Satisfactory) rating will normally only be given if the report is judged to be satisfactory in all those areas.

In some cases, the assessors may note in the Comments section that the rating is borderline, indicating that it is a matter of fine judgement as to whether the rating falls into one category or another. This most often happens when the assessors are deciding between B or C ratings.
The Glossary in Annex 1 defines many of the terms used in this Proforma.

---

[1] Sources used in the development of the Proforma are listed at the end of this document.
[2] The *Annual Review* series was renamed *Review of Humanitarian Action* series in 2004.
[3] Two assessors are used for the meta-evaluation exercise to mitigate potential assessor bias

EVALUATION TITLE	
COMMISSIONING AGENCY	
DATE OF REPORT	
NAME AND POSITION OF ASSESSOR	
REASON FOR ASSESSMENT	
DATE OF ASSESSMENT	
DATE OF AGENCY INTERVIEW (if held)	

Section 1. Assessing the Terms of Reference (ToR)

Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
1.1 The Terms of Reference	The ToR should clearly describe: (a) The work to be evaluated including its objectives and key stakeholders. *(b)* The purpose, objectives and focus of the evaluation *(Purpose might be donor requirement, accountability, lesson learning, community empowerment. Focus might be on partner performance, programme, project, policy, institutional analysis, sector, coordination).* (c) The intended use and users of the evaluation outputs and the individual or department responsible for follow-up. (d) The desired report framework. (*A sample framework is outlined in Annex 2).* (e) The rationale for the timing of the evaluation. (f) The evaluator selection process (e.g., competitive bidding, standing offer).		
1.2 Expectation of good evaluation practice	The TOR should clarify the commissioning agency's expectation of good humanitarian evaluation practice. *(e.g., application of DAC criteria;*[4] *reference to international standards including international law; multi-method approach i.e., quantitative and qualitative; consultation with key stakeholders to inform findings, conclusions and recommendations; and gender analysis).*		

[4] See Section 5.3 below for criteria definitions drawn from OECD/DAC (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, Paris.

Section 2. Assessing Evaluation Methods, Practice and Constraints

Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
2.1 Nature, make up and appropriateness and biases of the evaluation team	a) The report should outline the nature (e.g., external or mixed) and make up of the team (e.g., sectoral expertise, local knowledge, gender balance) and its appropriateness for the evaluation. b) The evaluation report should outline the evaluator(s)' biases that might have affected the evaluation and how these have been counteracted.		
2.2 Clarification process	The evaluation report should outline any clarification process between the commissioning agency and the evaluation team about the scope and methods of the evaluation that resulted in modifications to the ToR.		
2.3 Appropriateness of the overall evaluation methods	The evaluation methods should be clearly outlined in the report and their appropriateness, relative to the evaluation's primary purpose, focus and users, should be explained pointing out the strengths and weaknesses of the methods.		
2.4 Consultation with and participation by primary stakeholders	(a) The evaluation report should outline the nature and scope of consultation with, and participation by, beneficiaries and non- beneficiaries within the affected population in the evaluation process. (A satisfactory or higher rating should only be given where evidence is presented of adequate consultation and participation of primary stakeholders in the evaluation process, or where, in the assessor's view, it has been successfully argued as inappropriate due to security or other reasons.) (b) The evaluation report should outline the nature and scope of consultation with other key stakeholders in the evaluation process. The report should include a list of the other key stakeholders who were consulted or who participated in the evaluation process.		
2.5 The use of and adherence to international standards	The evaluation report should assess the intervention against appropriate international standards (e.g., international humanitarian and human rights law; the Red Cross/ NGO Code of Conduct, Sphere).		
2.6 Evaluation constraints	The evaluation report should outline key constraints to carrying out the evaluation (e.g., lack of time, difficult travelling conditions, lack of baseline data, poor agency monitoring systems, lack of access to key information sources, difficulties setting up control groups, use of translators), and the effect of these constraints.		

Section 3. Assessing Contextual Analysis

Area of enquiry	Guidance Notes	Comments	Rating
3.1 Analysis of context and of the crisis to which the intervention is responding	(a) The evaluation report should provide analysis of the affected area and population *(including relevant historical, social, economic, political and cultural factors)* to inform the evaluation and draw on this information in the text to support the analysis of the intervention. (b) The evaluation report should provide a clear analysis of the crisis, including key events (and a chronology where appropriate).		
3.2 Past involvement of the agency and its local partners	The evaluation report should provide analysis of the implementing agency's and its local partners' past involvement and main areas of work, so that the influence of the agency's past involvement on the intervention, including its geographical and sectoral focus, can be understood.		

Section 4. Assessing the Intervention

4.1 Institutional Considerations			
Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
4.1.i The agency's guiding policies and principles	The evaluation report should provide an analysis of the extent to which agency policies and principles were applied, and their relevance to and effect on the intervention.		
4.1.ii The agency's management and human resources	The evaluation report should provide an analysis of the agency's management and human resource procedures and practices as applied and their effect on the intervention. *(This might include: level of experience/expertise of field staff; use of national and expatriate staff; staff turnover; field/HQ communications & relations; briefing and debriefing procedures; training and learning practices; security)*		

4.2 Needs Assessment, Objectives, Planning and Implementation			
Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
4.2.i The needs and livelihoods assessments that informed the intervention	The evaluation report should provide analysis of the needs and livelihoods assessment practices that informed the intervention and their effect on the intervention.		
4.2.ii Intervention objectives	The evaluation report should assess the relevance of the intervention objectives to the contextual analysis and needs/livelihoods assessments assessed in 3.1 and 4.2.i above.		

4.2.iii Programme cycle processes.	The evaluation report should provide analysis of the following processes and their effect on the intervention: (a) planning (b) implementation (c) monitoring and/or real-time evaluative mechanisms (d) intervention expenditure. *(Consideration in this analysis should be given to local capacities; primary stakeholder consultation and participation; local and national partnerships)*		

4.3 Application of EHA Criteria			
Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
	The evaluation report should provide evidence of an adequate application of standard evaluation of humanitarian action criteria as per the OECD/DAC definitions given below:[5]		
4.3.i Efficiency (including cost-effectiveness)	Efficiency measures the outputs - qualitative and quantitative - in relation to the inputs. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been used. Cost-effectiveness looks beyond how inputs were converted into outputs, to whether different outputs could have been produced that would have had a greater impact in achieving the project purpose.		
4.3.ii Effectiveness (including timeliness)	Effectiveness measures the extent to which the activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criteria of effectiveness is **timeliness** of the response. (*Although **coordination** is not a formal criterion, the OECD/DAC Guidance suggests that given its importance, it should be considered under this criterion).*		

[5] from OECD/DAC (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies.* Paris, pp 30-32.

4.3.iii Impact	Impact looks at the wider effects of the project - social, economic, technical, environmental - on individuals, gender, age-groups, communities, and institutions.		
4.3.iv Relevance/ appropriateness	Relevance is concerned with assessing whether the project is in line with local needs and priorities (as well as donor policy). It refers to the overall goal and purpose of a programme. Appropriateness - the need to tailor humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly … is more focused on the activities and inputs.[6]		
4.3.v Sustainability/ connectedness	Sustainability is concerned with measuring whether an activity or an impact is likely to continue after donor funding has been withdrawn. … many humanitarian interventions, in contrast to development projects, are not designed to be sustainable. They still need assessing, however, in regard to whether, in responding to acute and immediate needs, they take the longer-term into account. (*Minear has referred to this as Connectedness, the need… to assure that activities of a short-term emergency nature are carried out in a context which takes longer-term and inter-connected problems into account.*[7])		
4.3.vi Coverage	The need to reach major population groups facing life-threatening suffering wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agendas.		
4.3.vii Coherence	Coherence refers to policy coherence, and the need to assess security, developmental, trade and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations.		

[6] Minear, L. (1994) The International Relief System: A critical review. Paper presented to the Parallel National Intelligence Estimate on Global National Emergencies, Meridian International Centre, Washington DC, September 2002.
[7] Minear, L. (1994) *ibid*.

4.4 Consideration given to Cross-cutting Issues			
Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
4.4.i The use of and adherence to international standards	The evaluation report should assess the extent to which relevant international standards were used in the planning, implementation and monitoring of the intervention *(e.g., international humanitarian and human rights law; the Red Cross/ NGO Code of Conduct and developing standards - e.g., Sphere)*		
4.4.ii Gender Equality	The evaluation report should analyse consideration given to gender equality throughout the intervention and the effect on the intervention. *(i.e. was gender equality taken into consideration in all relevant areas? Did the intervention conform to the implementing organisation's gender equality policy? It should be noted if there is no gender equality policy).*		
4.4.iii Protection	The evaluation report should analyse the consideration given to protection throughout the intervention cycle and the effect on the intervention.		
4.4.iv Capacity building	The evaluation report should analyse the consideration given to the capacity building of key and primary stakeholders government and civil society institutions, and the effect of this on the intervention.		
4.4.v Advocacy	The evaluation report should analyse consideration given to advocacy and the effect on the intervention. *(e.g., attempts to influence donors, partners, government, concerning their policies or actions).*		
4.4.vi Vulnerable and marginalised groups	The evaluation report should provide an analysis of consideration given to vulnerable and marginalised groups (e.g., elderly, disabled, children, HIV/AIDS sufferers) and to other groups that suffer discrimination and disadvantage.		

Section 5. Assessing the Report

5.1 Findings, Conclusions and Recommendations			
Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
5.1.i Secondary sources	The evaluation report should use and refer to relevant secondary sources to support its findings, conclusions and recommendations (*a satisfactory or higher rating should only be given where a reference list of secondary sources is included as part of the report).*		
5.1.ii Conclusions	The report's conclusions should flow logically from, and reflect, the report's central findings. The report should provide a clear and defensible basis for value judgements in each case.		
5.1.iii Recommendations	(a) Recommendations should be clear, relevant and implementable, reflecting any constraints to follow up. (b) Recommendations should follow on from the main conclusions and reflect consultation with key stakeholders. (c) The evaluation report should suggest a prioritisation of recommendations, timeframe for implementation and suggest where responsibility for follow-up should lie if that is not indicated in the ToR.		

5.2 Report Coverage, Legibility and Accessibility			
Area of enquiry	**Guidance Notes**	**Comments**	**Rating**
5.2.i Coverage of the evaluation report	The evaluation report should adequately cover all areas specified in the ToR and additional factors that affected the performance of the intervention.		
5.2.ii Format of the report	The evaluation report format should follow that outlined in the ToR (*if the ToR did not propose a format for the report, this area should be assessed on the basis of the good practice suggested in Annex 2).*		
5.2.iii Accessibility of the report	The evaluation report should cater for the intended readership and users (*In general reports should use language clearly; be succinct; be clearly laid out e.g. with different information levels and appropriate visual aids. Some organisations have their own style guides).*		
5.2.iv Executive Summary	The executive summary should reflect the format of the main text, and clearly outline key evaluation conclusions and recommendations.		

Section 6. Overall Comments *(for information purposes and not rated)*

Area of enquiry	Guidance Notes	Comments
6.i Comments on issues not covered above.	This is an opportunity for comment on any issues not covered by the areas of enquiry.	
6.ii Overall comments on the report.	This is an opportunity to make an overall comment on the report, including its strengths and weaknesses.	

Annex 1

GLOSSARY

Accountability
Accountability is the means by which individuals and organisations report to a recognised authority, or authorities, and are held responsible for their actions. (Edwards & Hulme, 1995).

Advocacy
Advocacy refers in a broad sense to efforts to promote, in the domain of humanitarian aid, respect for humanitarian principles and law with a view to influencing the relevant political authorities, whether recognised governments, insurgent groups or other non-state actors. (Swiss Agency for Development and Cooperation, 2004).[8] One could add "international, national and local assistance agencies".

Appropriateness
The need to "tailor humanitarian activities to local needs, increasing ownership, accountability, and cost- effectiveness accordingly" (Minear 1994) … is more focused on the activities and inputs.[9] (ALNAP Annual Review Glossary 2003)

Coherence
Refers to the policy coherence and the need to assess security, developmental, trade and military policies to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations. (DAC Evaluation Criteria)

Complex political emergency
A situation with complex social, political and economic origins which involves the breakdown of state structures, the disputed legitimacy of host authorities, the abuse of human rights and possibly armed conflict, that creates humanitarian needs. The term is generally used to differentiate humanitarian needs arising from conflict and instability from those that arise from natural disasters. (ALNAP Annual Review Glossary 2003)

Conclusions
Conclusions point out the factors of success and failure of the evaluated intervention, with special attention paid to the intended and unintended results and impacts, and more generally to any other strength or weakness. A conclusion draws on data collection and analyses undertaken through a transparent chain of arguments. (OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management, 2002)

Context (of an evaluation)
The combination of factors accompanying the study that may have influenced its results, including geographic location, timing, political and social climate, economic conditions, and other relevant professional activities in progress at the same time. (Programme Policy and Procedures Manual, UNICEF, May 2003)

---

[8] NB Definitions of advocacy within the humanitarian sector appear to be very limited. Swiss Agency for Development and Cooperation. 2004. *Advocacy Guidelines: Humanitarian Aid of the Swiss Confederation.* Berne. March 2004.
[9] Minear, L. (1994) The International Relief System: A critical review. Paper presented to the Parallel National Intelligence Estimate on Global National Emergencies, Meridian International Centre, Washington DC, September 2002.

Cost Effectiveness Analysis (see also 4.3.i above)
Cost-effectiveness analysis entails comparing costs across different strategies for achieving a given outcome, with a view to determining the lowest cost approach. For example, cost-effectiveness analysis might explore three different approaches to getting girls working in the informal sector back into school. As compared to cost-efficiency analysis, it is wider in scope, looking beyond outputs to outcomes. (M&E Training Resources, UNICEF, 2004)

Coverage
The need to reach major population groups facing life-threatening suffering wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agenda. (DAC Evaluation Criteria)

Effectiveness
Effectiveness measures the extent to which the activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criteria of effectiveness is **timeliness** of the response. Although **coordination** is not a formal criterion, the OECD/DAC Guidance suggests that given its importance, it should be considered under this criterion. (DAC Evaluation Criteria)

Humanitarian action
Assistance, protection and advocacy actions undertaken on an impartial basis in response to human needs resulting from complex political emergencies and natural hazards. (ALNAP Annual Review Glossary 2003)

Impact
Impact looks at the wider effects of the project - social, economic, technical, environmental - on individuals, gender, age-groups, communities, and institutions. (DAC Evaluation Criteria)

Impartiality
An approach to the provision of humanitarian assistance and services which is non-discriminatory, proportionate to needs and free of subjective distinction. A guiding principle of organisations claiming to be humanitarian. (ALNAP Annual Review Glossary 2003)

Input
The financial, human, material, technological and information resources used for the intervention. (OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management Proposed Harmonized Terminology, 2002)

Lesson learned
Conclusions that can be generalized beyond the specific case. This could include lessons that are of relevance more broadly within the country situation or globally, to an organization or the broader international community. (Programme Policy and Procedures Manual, UNICEF, May 2003)

Lesson-learning study A study initiated by an organisation with the explicit objective of lesson-learning within that organisation, but that falls outside the full evaluation definition. A process that may be facilitated by external consultants but is generally an internal process.(ALNAP Annual Review Glossary 2003)

Meta-evaluation Simply stated, meta-evaluation is the evaluation of an evaluation, evaluation system or evaluation device (Hummel 2003). A process of delineating, obtaining, and applying descriptive information and judgmental information – about the utility, feasibility, propriety and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness and social responsibility – to guide the evaluation and/or report its strengths and weaknesses (Stufflebeam)

Outcome
The intended or achieved short-term and medium-term effects of an intervention's outputs, usually requiring the collective effort of partners. Outcomes represent changes in conditions which occur between the completion of outputs and the achievement of impact. (OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management Proposed Harmonized Terminology, 2002)

Output
The products and services which result from the completion of activities within an intervention. (OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management Proposed Harmonized Terminology, 2002)

Protection
Activities aimed at obtaining full respect for the rights of the individual in accordance with the letter and the spirit of the relevant bodies of law (i.e. human rights, humanitarian and refugee law) which are conducted impartially and not on the basis of race, national or ethnic origin, language or gender. (ALNAP Annual Review Glossary, 2003)

Relevance
Relevance is concerned with assessing whether the project is in line with local needs and priorities (as well as donor policy) … refers to the overall goal and purpose of a programme. (DAC Evaluation Criteria)

Retrospectively, the question of relevance often becomes a question as to whether the objectives of an intervention or its design are still appropriate given changed circumstances. (OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management, 2002)

Stakeholder
All those – from agencies to individuals – who have a direct or indirect interest in the humanitarian intervention, or who affect or are affected by the implementation and outcome of it. (ALNAP Annual Review Glossary 2003). Within the context of the Quality Proforma **'primary stakeholders'** refers to both beneficiaries and non-beneficiaries within the affected population.

Sustainability
Sustainability 'is concerned with measuring whether an activity or an impact is likely to continue after donor funding has been withdrawn … many humanitarian interventions, in contrast to development projects, are not designed to be sustainable. They still need assessing, however, in regard to whether, in responding to acute and immediate needs, they take the longer term into account. (DAC Evaluation Criteria). Minear has referred to this as *Connectedness. Connectedness*, the need "to assure that activities of a short term emergency nature are carried out in a context which takes longer-term and inter-connected problems into account" (Minear, 1994).

Terms of Reference
Terms of reference define the requirements and paramters for conducting an evaluation. (ALNAP Annual Review Glossary 2003)

Annex 2

EVALUATION REPORT FORMAT - CHECK LIST

Preliminaries	
	Title page (should include date of report)
	List of contents with page numbers
	Acronyms
	Map(s)
	Executive Summary
Main text	
	Introduction (including motivation for commissioning evaluation, purpose of study, scope, approach, methods, composition of team, constraints)
	Context in which humanitarian action took place, humanitarian context and response
	Findings
	Conclusions
	Recommendations
Annexes	
	Sources/bibliography
	ToR
	Timetable
	Evaluation team profiles
	List of Interviewees
	Timeline
	Evaluation Material (questionnaires etc.)
	Collated stakeholder feedback on findings, conclusions and recommendations
	Other appendices/annexes

References

Apthorpe, R. (2000) K*osovo Humanitarian Programme Evaluations: Towards Synthesis, Meta-Analysis and Sixteen Propositions for Discussion.* Background Paper for the ALNAP Symposium 'Learning from Evaluation: Humanitarian Assistance and Protection in Kosovo' 17th & 18th October 2000, Geneva (London:ALNAP)

Patton, M. (1997) *Utilization-Focused Evaluation: The New Century Text* (Thousand Oaks: Sage) Raynard, P. (2000)

Mapping Accountability in Humanitarian Assistance (London: ALNAP)

SPHERE (2000) *Humanitarian Charter and Minimum Standards in Disaster Response* (Geneva: The Sphere Project)

Valadez, J. and M. Bamberger (1994) *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers and Researchers* (Washington DC: World Bank EDI Development Series)

Wood, A., Borton, J. and R. Apthorpe (eds.) (2001) *Evaluating International Humanitarian Action: Reflections from Practitioners* (London: Zed Press)

Annex 2 WASH Accountability Checklist

ASK, LISTEN, COMMUNICATE WASH Accountability Checklist ASK, LISTEN,

"We acknowledge that our fundamental accountability must be to those we seek to assist."

Sphere Humanitarian Charter

Accountability can be seen as having 5 dimensions:

I. **Transparency T**
II. **Feedback and Complaints F**
III. **Participation P**
IV. **Staff competencies and attitudes S**
V. **Monitoring and Evaluation M**

In practical terms the emphasis is on **active listening, a respectful attitude and the ability to empathise with those affected.** The formal and informal **provision of adequate and timely information** is also critical. The checklist is not prescriptive and it is intended that it will be agreed by each team and adapted to their specific context.

	First Phase/Acute	Dimension	Scale 1-5
1.	Agency staff members always **treat affected women, men and children with respect** (e.g. showing courtesy and patience, meeting them on their terms, actively seeking their viewpoint on all programme issues, sharing project information openly and recognising the vital contribution this has to people's recovery). HAP Benchmark 4	S	
2.	Agency staff members **do not abuse their position of power as per the Staff Code of Conduct** (e.g. by asking for any kind of favour from those affected in return for assistance). This is a serious offence and provides grounds for disciplinary proceedings. HAP Benchmark 4	S	
3.	WASH personnel do not enter a community / camp / settlement without first discussing and **seeking permission** from the local leadership or representatives where available (but they should be aware of the extent to which such leaders may or may not represent all community members). HAP Benchmark 3	P	
4.	**Photographs must not be taken indiscriminately** and wherever possible the consent of the person being photographed must be obtained. HAP Benchmark 2 & 3	T, P	
5.	Where vehicles are used, they are **parked discreetly** and the WASH team does not consult with community members from their vehicles. **Sit at the same level** as the person you are discussing with e.g. on the ground or crouching down. HAP Benchmark 3	P	
6.	WASH staff will **dress appropriately** and take cultural norms into consideration (e.g. covering legs and arms, not wearing tight trousers, removing sunglasses when talking to people etc.) HAP Benchmark 4	S	
7.	WASH personnel will **plan community meetings** in advance at a time that is mutually convenient and allows maximum participation. They will attend all meetings on time or provide an explanation for the delay (which may be due to late arrival of community members). HAP Benchmark 3	P	
8.	The **recruitment of both male and female staff must be ensured** and female staff especially should be available to consult with female community members. HAP Benchmark 3 & 4	P, S	
9.	Where WASH personnel do not speak the same language as the affected community, **interpreters must be provided.** WASH personnel should also limit discussions amongst themselves that cannot be understood by the affected community. Every effort must be made to **recruit WASH personnel who speak the same language as those affected.** HAP Benchmark 2 & 4	T, S	
10.	All WASH personnel **will identify themselves** (e.g. by using badges stating name and agency/wearing ID cards with photographs) when working in the field. (Security considerations must be adhered to and take precedence). HAP Benchmark 2	T	
11.	Agency staff members will **consult women and men separately** and will discuss the programme with **the most vulnerable groups (e.g. children, older people and those with disabilities)**. They will identify the main social groupings in the community, including the most marginalised, and consider their priorities. HAP Benchmark 3	P	
12.	People/communities are **systematically provided with information** on the following: • Name of organisation and contact details • Names of staff working in their location • Proposed plan of action / follow up visits / feedback • Source of funding and proposed expenditure (where this does not compromise security) • Type and duration of intervention • Dates and times for community meetings • Targeting criteria (should be agreed previously through community discussions) • How to make a complaint and the right to receive a reply • Agreed roles & responsibilities of agency and communities • Expected standards of behaviour HAP Benchmark 2	T	
13.	People/communities are informed that they have the **right to provide feedback or suggestions** (good or bad) and a right to receive a reply on the work being undertaken and on staff conduct. HAP Benchmark 2	F, P, M	
14.	Staff must respond to all complaints received in a timely manner. HAP Benchmark 2	F	

ASK, LISTEN, COMMUNICATE WASH Accountability Checklist ASK, LISTEN,

15.	Feedback on the **design and siting/location** of toilets, washrooms, laundry slabs and water points will be actively sought from women, men, boys and girls as well as people with special needs or vulnerabilities such as those with disabilities. HAP Benchmark 3	F, P, M	
16.	Discussions must be held, with **women, children, older people and those with disabilities** on **safe access** to water & sanitation facilities e.g. location of facilities, provision of locks, lighting, clearing of bushes etc. HAP Benchmark 3	P	
17.	The **degree of satisfaction** with first phase hygiene kits or NFI's will be monitored as quickly as possible following distribution and **input sought into the content** of subsequent hygiene kits (and preferred brands where feasible). HAP Benchmark 3	F, P, M	
18.	Consultation **with women by women** will be held to determine appropriate requirements for female menstruation and the method of provision (e.g. lingerie fairs, vouchers etc). HAP Benchmark 3	P	
19.	WASH staff will **identify the local institutions** responsible for delivering water and sanitation services, discuss plans with them and **seek their input and guidance.** HAP Benchmark 3	P	
20.	**Communication, collaboration and co-ordination with other WASH staff and other WASH agencies** should be seen as a key aspect of every fieldworker's job description., HAP Benchmark 4	S	
21.	The decision to pay for community labour or outreach systems must be considered in the light of its impact on longer-term development and **where unavoidable should be explained** to the affected population. HAP Benchmark 2 & 3	T, P	
22.	**Monitoring systems** should seek to involve community members and to identify their perspective on success and failure as well as measure client satisfaction and degree of participation. HAP Benchmark 3 & 6	P, M	
23.	Feedback on all **HP communication materials** should also be obtained and this should inform future adaptation and development HAP Benchmark 3 & 6	P, M	
24.	The WASH team will regularly assess **what social groups have been consulted and which omitted** and attempts will be made to ensure ongoing contact with the most vulnerable groups. HAP Benchmark 3 & 6	P, M	
25.	Where feasible **relevant feedback** from assessments, monitoring and evaluation **will be provided to community** members (and other stakeholders including government) and they should have an opportunity to comment on the findings and process of the assessment or monitoring. HAP Benchmark 3	T, P	
26.	**ALL WASH staff** (including community volunteers or mobilisers) **will be adequately prepared** to enable them to work in a transparent and supportive way with those affected (recruitment, job descriptions, orientation, training, ongoing mentoring and support). HAP Benchmark 4	S	
Second Phase/Chronic Emergency		Dimension	Scale 1-5
1.	**MoUs or partnership agreements** should be signed with user groups to ensure that the respective responsibilities of service provider and user are adhered to. HAP Benchmark 3	T, P	
2.	Attention must be paid to **long-term management** of facilities by strengthening links with, and the capacity of, national structures and systems/user groups/committees etc. HAP Benchmark 1	P, M	
3.	User groups should be provided with **the appropriate capacity to ensure their long-term functioning** e.g. technical training and training in group management, leadership, conflict mitigation, budgeting etc. HAP Benchmark 3	P	
4.	**Ongoing information** is systematically provided to affected women and men using the means that they favour (e.g. community meetings, information sheets, notice boards with pictures, radio, posters, newspapers etc) and the language of their choice. HAP Benchmark 2 & 3	T	
5.	Programmes will be designed **according to each specific context, based on discussion with users** and a standard 'one size fits all' approach will be avoided. Management options must also be discussed. HAP Benchmark 2	T	
6.	WASH facilities must be **completed as agreed** and should enable communities to maintain them easily (e.g. plastering, drainage, etc.) HAP Benchmark 1	M	
7.	Organisational branding (use of logos on shelters, NFIs, notice boards etc.) will not be used indiscriminately without the consent of affected populations. HAP Benchmark 2 & 3	T, P	
8.	Where feasible agency staff members should try to make some **time for informal, discussions** during the project cycle - discussing whatever is important to local people. HAP Benchmark 3	P	
9.	**Regular reports of actual performance** in relation to previously agreed goals are provided to the affected communities. Those affected should have an opportunity to ask questions, discuss and/or challenge these. HAP Benchmark 2 & 3	T	
10.	Any **significant changes** to programme goals or activities; or the budget; or to key contacts are provided promptly to representatives/user groups. HAP Benchmark 2	T	
11.	**Ways should be sought to provide information on expenditure** e.g. the costs of materials/suppliers etc. and how community members can help to guard against corruption. HAP Benchmark 2 & 3	T, P	
12.	**Managers model open, inclusive and respectful behaviour** within the staff team (e.g. making decisions in a transparent and inclusive way, welcoming divergent views and encouraging collaboration and team-working) HAP Benchmark 1 & 4	P, S	
13.	All staff can explain the meaning of 'accountability to the community' and can give examples of how they have been accountable to the people they provide services to. HAP Benchmark 4	S	
14.	**A robust quality management system** must be put in place to support staff and project development. HAP Benchmark 1	All	
15.	**A clear and transparent 'exit strategy'** must be discussed with and communicated to affected	T, P	

ASK, LISTEN, COMMUNICATE WASH Accountability Checklist ASK, LISTEN,			
	communities. **HAP Benchmark 3** **(ADD OR ADAPT INDICATORS AS REQUIRED)**		

Adapted from: Accountability to Beneficiaries: A Practical Checklist Mango 2007

Resources:

The Good Enough Guide: Impact Measurement and Accountability in Emergencies (2007) Emergency Capacity Building Project:
http://www.oxfam.org.uk/what_we_do/resources/downloads/Good_Enough_Guide.pdf

Sphere: Humanitarian Charter and Minimum Standards in Disaster Response (Includes **the Red Cross Code of Conduct**): www.sphereproject.org

HAP International (2008): The guide to the HAP Standard, Humanitarian Accountability and Quality Management, available online at
http://publications.oxfam.org.uk/oxfam/add_info_051.asp

HAP 2007 Standard in Humanitarian Accountability and Quality Management, available online at http://www.hapinternational.org/standards.aspx

Abstract

As disasters are complex, unique and dynamic phenomena, and evaluations are not a tool to capture everything, the lenses through which evaluations look at things are critical. The evaluation criteria developed by the *Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD-DAC)* have become prevalent in the *evaluation of humanitarian action (EHA)*. The research of this diploma thesis explores the OECD-DAC criteria for evaluating *Water, Sanitation and Hygiene Promotion (WASH)* interventions in humanitarian response to disasters. In the focus are their recent practice, what they are used to reveal and if that is addressing persistent issues in the sector to create an evidence base for learning. The discussion of the OECD-DAC criteria's usefulness for evaluating WASH emergency interventions is framed with three research questions: *How are the OECD-DAC criteria used to evaluate humanitarian emergency response in the WASH sector and what do they reveal? How are cross-cutting themes and standards used in conjunction with the OECD-DAC criteria? What are the benefits and shortcomings of using OECD-DAC criteria for evaluating WASH emergency response activities?*

At the heart of the research is a meta-analysis of evaluation reports. It is complemented by literature reviews, documentary research and a review of existing EHA approaches. The thesis argues that because theory of change models are hardly applicable in emergency response, deriving lessons from field experience gained through trial and error is inevitable. Accordingly, the best possible utilisation of evaluations would be to capture this process of testing. However, current evaluations hardly fulfil this purpose, particularly with respect to the most prominent and persistent weakness of WASH emergency interventions, i.e. linking relief and development. The research demonstrates that the OECD-DAC criteria largely reveal what is already well known and emphasised in humanitarian standards developed by the *Sphere Project* and the *Humanitarian Accountability Partnership*. While some of the OECD-DAC criteria are irreplaceable for measuring the performance of WASH interventions, using established humanitarian standards is the only way to assess relevant quality and accountability aspects. It is concluded that a focused use of a few selected OECD-DAC criteria (according to the evaluation purpose) paired with humanitarian standards provide a more fruitful basis for evidence on *What works, what not and why?* Additionally, such an approach would aid on-going attempts to make evaluations less intricate endeavours, eventually leading to an active role of disaster-affected people in evaluations.

Kurzfassung

Katastrophen sind komplexe, einzigartige und dynamische Phänomene. Da Evaluierungen nicht alle Aspekte einer solchen erfassen können, ist es entscheidend, wie die zu evaluierenden Fragestellungen analysiert werden. Die vom *Development Assistance Committee* der *Organisation for Economic Co-operation and Development (OECD-DAC)* entwickelten Kriterien sind in der Evaluierung humanitärer Hilfe weit verbreitet. Die Diplomarbeit beschäftigt sich mit der Anwendung der OECD-DAC Kriterien zur Evaluierung von *Water, Sanitation and Hygiene Promotion (WASH)* Interventionen in der humanitären Katastrophenhilfe. Der Forschungsschwerpunkt liegt dabei auf ihrer Anwendungspraxis, ihrem Beitrag zum Erkenntnisgewinn und inwiefern sie dazu geeignet sind, zentrale WASH Themen zu beleuchten, um evidenzbasiertes Lernen zu ermöglichen. Die Diskussion über die Eignung der OECD-DAC Kriterien für humanitäre Katastrophenhilfe im Bereich WASH bezieht sich auf drei Forschungsfragen: *Wie werden die OECD-DAC Kriterien angewendet um humanitäre WASH Katastropheninterventionen zu evaluieren und was enthüllen sie? Wie werden Querschnittsthemen und Standards in Verbindung mit den OECD-DAC Kriterien angewendet? Welche Stärken und Schwächen lassen sich für die Verwendung der OECD-DAC Kriterien in der Evaluierung von Aktivitäten humanitärer Katastropheneinsätze ausmachen?*

Den Kern der Untersuchung bildet eine Metaanalyse von Evaluierungsberichten. Die Forschungsmethodologie wird durch eine Kombination aus Literatur- und Dokumentenrecherche sowie einer Analyse bestehender Evaluierungsansätze ergänzt. Die Forschungsarbeit argumentiert, dass theoretische Veränderungsmodelle in der Katastrophenhilfe schwer anwendbar sind und es daher umso wesentlicher ist, Schlüsse aus Erfahrungen systematischen Testens verschiedener WASH Ansätze zu ziehen. Demnach wäre das Einfangen dieser Testerfahrungen die am besten geeignetste Anwendung von Evaluierungen. Allerdings erfüllen Evaluierungen diesen Zweck bisher nur in geringem Ausmaß, insbesondere in Bezug auf die bedeutendste Schwäche von WASH Katastropheninterventionen: Die Verbindung von Nothilfe mit längerfristiger Entwicklung. Die Forschungsarbeit zeigt, dass die OECD-DAC Kriterien vor allem Themen adressieren, welche in den vom *Sphere Project* und *Humanitarian Accountability Partnership* entwickelten humanitären Standards bereits Ausdruck finden. Während einige der OECD-DAC Kriterien zur Bestimmung der Performance von WASH Interventionen unersetzbar sind, sind die qualitativen Aspekte solcher nur mittels etablierter humanitärer Standards bestimmbar. Aus den Forschungsergebnissen folgt, dass ein fokussierter Einsatz einiger weniger OECD-DAC Kriterien (gemäß dem Evaluierungszweck), verbunden mit humanitären Standards, besser geeignet ist, um zu ergründen welche WASH Konzepte funktionieren, welche nicht und warum. Zusätzlich würde ein solcher Ansatz bestehende Bestrebungen unterstützen Evaluierungen einfacher zu machen wie letztlich den von Katastrophen betroffenen Menschen eine aktive Rolle im Evaluierungsprozess zu ermöglichen.

Curriculum Vitae

Philipp Polanski
(*1985, Steyr, Austria)

Contact
philipp.polanski@gmx.at

Work experience

since 09/2012	Export Clerk (Logistics Company)
06/2013 – 07/2013	WASH Technician, Austrian Red Cross (Emergency Drinking Water Supply for Flood-affected Communities in Upper Austria)
03/2012 – 04/2012	WASH Delegate, Austrian Red Cross (WASH Training and Emergency Response in Kosti and Al Salam, Sudan)
03/2011 – 03/2011	WASH Delegate, Austrian Red Cross (WASH Disaster Response Kit Trainings in Khartoum, Sudan)
08/2010 – 10/2010	Internship at the *National Housing Strategy Project for the Kingdom of Saudi Arabia* of the General Housing Authority (GIZ International Services in Riyadh, Saudi Arabia)
02/2010 – 03/2010	WASH Delegate, Austrian Red Cross (IFRC Floods Operation in Dodoma/Morogoro Region, Tanzania)
10/2009 – 01/2010	Warehouse Clerk (Wholesale Distribution Business)
09/2003 – 09/2004	Paramedic (Community Service)
Summer 2000 – 2009	Labourer (Sawmill; Construction Company; Truck Manufacturer; Glass Producer)

Education

since 10/2004	University of Vienna: *Development Studies*
	General: Sustainable and Human Rights-based Development Occupational: Services of General Interest; Social and Economic Security
10/2004 – 11/2012	University of Vienna: Diploma Degree in *Geography* (Mag.rer.nat., equivalent to Master)
	General: Applied Geography – Spatial Research and Planning Occupational: Geographic Information Systems; Natural Resources Management
09/1995 – 06/2003	Secondary School for General Education (Bundesgymnasium Werndlpark in Steyr)

Languages

German (mother tongue), English and Spanish

Miscellaneous

since 04/2006	Volunteer of the Austrian Red Cross as Emergency Response Unit *Water and Sanitation* Roster Member, and Trainer for WASH and Geographic Information Systems in Emergencies
10/2004 – 12/2009	Volunteer of the Austrian Red Cross as Paramedic

Herewith I affirm:

- That this diploma thesis was written by myself.
  I did not use any sources besides those cited as references.
- That this diploma thesis was not submitted prior to this as an examination paper in any form.
- That this thesis is identically equal to that evaluated by the adviser.

Ich versichere:

- Dass ich die Diplomarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- Dass ich dieses Diplomarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.
- Dass diese Arbeit mit der von der Begutachterin beurteilten Arbeit übereinstimmt.

Date/Datum　　　　　　　　　　　　　　　　　　　　　　Signature/Unterschrift