

**INVESTIGATING EVALUATION AS AN ACCOUNTABILITY MECHANISM BY
INTERNATIONAL NON-GOVERNMENTAL ORGANIZATIONS WORKING IN
HUMANITARIAN RELIEF**

by

Christopher Morris

M.A., The University of Denver, 2003

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Measurement, Evaluation and Research Methodology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2014

© Christopher Morris, 2014

Abstract

Accountability of humanitarian relief organizations has been a key topic of discussion since the Joint Evaluation of Emergency Assistance to Rwanda was published in 1996. Dozens of initiatives stressing accountability to beneficiaries have been launched. However, humanitarian organizations still receive criticism for focusing on accountability to donors and ignoring their responsibilities to account for their actions to the communities they serve. Evaluation is considered a key mechanism for providing accountability and can give opportunities for reducing power imbalances. This study investigates how humanitarian International Non-Governmental Organizations (INGOs) are using evaluation, asking whether evaluation practice is providing accountability to communities affected by crisis. Using a critical hermeneutic framework, the study undertook an empirical review of a sample of evaluation reports published on the Active Learning Network for Accountability and Performance (ALNAP) website. Interviews with evaluators and INGO staff involved in the evaluations contributed to the understanding of current evaluation practice. The study found the accountability provided was mainly internal to INGOs and accountability to affected communities was low. Ensuring program improvement through evaluation was a weak form of accountability but affected communities were not able to use evaluation to influence decisions that affect them. Participation in evaluations was limited to the inclusion of beneficiaries at the data collection stage, and there was no evidence of participation in developing the evaluation scope or questions. Participation at the final stages of the evaluation was also low, although the evaluations that included local civil society partners showed evidence of community involvement in either negotiating or receiving the evaluation results. These latter evaluations provided the highest degree of accountability to the community. Opportunities for participatory

evaluation approaches were constrained by INGO control of the evaluation scope and the time allocated for the evaluation. As a result evaluation approaches that favoured internal utilization rather than community engagement or empowerment were most common and thus INGOs benefitted the most from current evaluation practice.

Preface

This dissertation is an original intellectual product of the author, C. Morris. This research was granted approval by the University of British Columbia Behavioural Research Ethics Board. It holds an Ethics Certificate from the UBC Behavioural Research Ethics Board with the Certificate Number H13-00456.

Table of Contents

Abstract.....	ii
Preface.....	iv
Table of Contents	v
List of Tables	x
List of Abbreviations	xi
Acknowledgements	xiii
Chapter 1: Introduction	1
Research Questions	4
Humanitarian Relief.....	6
Research Assumptions	8
Contribution to the Body of Research	9
Chapter 2: Literature Review.....	11
NGO Accountability	11
The Evolution of Accountability Within the Humanitarian Sector	14
Disaster Responses and Accountability	14
Tracing Accountability Through the Humanitarian Accountability Reports	17
Summary	19
Accountability Organizations and Codes of Conduct.....	20
The Red Cross Code of Conduct	20
The Sphere Standards	20
ALNAP	21
People in Aid.....	21

The Humanitarian Accountability Partnership	21
Power and Accountability.....	22
Donor Governments.....	26
Host Governments	29
INGOs	31
Populations Affected By Crisis.....	33
Citizens of Donor Countries	35
Private Donors	36
Summary	39
New Public Management’s Influence on Accountability	40
Evaluation Within the Humanitarian Sector	43
EHA and Accountability.....	46
Summary	48
Evaluation Approaches	49
Evaluation Approaches Favouring Upward Accountability	49
Experimental and quasi-experimental designs.....	50
Theory based approaches.....	50
Impact evaluation.....	51
Technical evaluations.....	51
Evaluation Approaches Favouring Downward Accountability	52
Participatory evaluation (PE).....	52
Empowerment evaluation.	53
Collaborative evaluation.....	54

Utilization-focused evaluation (UFE).....	54
Deliberative democratic and democratic evaluation.....	55
Real-time evaluation (RTE).....	56
The Quality and Intent of Evaluation.....	57
The Logic of Evaluation	57
Chapter 3: Research Methodology.....	60
Research Questions	60
Researcher Background	61
Critical Hermeneutics	62
Summary	66
Data Collection Methods	67
Evaluation Reports.....	67
Sample selection.	67
Reading of evaluation reports.....	71
Interviews.....	72
Data Analysis	73
Chapter 4: Results.....	76
Evaluation Policies and Accountability Approaches.....	77
Evaluation Policies.....	77
Accountability Statements	83
Affected Community Involvement in the Evaluation.....	89
Terms of Reference and Evaluation Planning	90
Collecting Evaluation Data	94

Negotiating and Presenting the Results	98
Evaluation Criteria.....	101
Analysis.....	107
INGOs	107
Donors.....	108
Public in Donor Countries	110
Affected Communities	111
Evaluation Approaches and Methodologies	113
Use	117
Methods.....	121
Valuing.....	127
Who Benefits From Current Evaluation Practice?.....	129
INGOs	130
Donors.....	131
Local Institutions	132
Affected Communities	132
Conclusion	135
Chapter 5: Discussion and Implications	136
How Are INGOs Operating in Humanitarian Relief Contexts Using Evaluations as a Mechanism to Provide and/or Assess Accountability to Affected Communities?	136
Accountability Through Lesson Learning and Program Improvement	137
Accountability Through Information Sharing and Participation	138
Civil Society Involvement	141

Assessing Downward Accountability in Project Implementation	142
Policy vs. Practice	143
The Findings As a Whole	144
Where Does Power Lie?	145
What Particular Evaluation Approaches Are Being Used With Which Forms of Accountability in Humanitarian Relief Work?	147
Structural Problems.....	149
Time.....	150
Where evaluations sit within the project cycle.....	151
Understanding participation.....	153
Evaluation products.....	155
Methodologies.....	157
The DAC criteria.....	159
Who Is Benefitting?	161
INGOs	161
Affected Communities	162
Concluding Thoughts	164
Limitations and Future Research	165
Limitations	165
Future Research Directions	167
References.....	169
Appendices.....	193
Appendix A: List of Evaluation Reports Included in the Research Sample	193

List of Tables

Table 3.1 Summary of INGO membership/signatures on global accountability initiatives	87
Table 3.2 How beneficiaries are involved in the evaluation process.....	91
Table 3.3 Details of beneficiary consultation in the data collection stage of evaluations	97
Table 3.4 Criteria used in evaluations.....	102
Table 3.5 Classification of evaluations on Alkin and Christie's evaluation theory tree.....	110

List of Abbreviations

- ACF..... Action Contre La Faim
- ALNAP..... Active Learning Network for Accountability and Performance
- AfrEA..... African Evaluation Association
- CAP..... Consolidated Appeals Process
- CBO..... Community Based Organization
- CIDA..... Canadian International Development Agency
- DAC..... Development Assistance Committee
- DARA..... Development Assistance Research Associates
- DEC..... Disasters Emergency Committee
- DfID..... Department for International Development
- DRC..... Danish Refugee Council
- EHA..... Evaluation of Humanitarian Action
- GHD..... Good Humanitarian Donorship
- GHIT..... Global Humanitarian Indicator Tool
- HAP..... Humanitarian Accountability Partnership
- HRI..... Humanitarian Response Index
- IASC..... Inter-Agency Standing Committee
- IATI..... International Aid Transparency Initiative
- ICBL..... International Campaign to Ban Landmines
- ICRC..... International Committee for the Red Cross
- IFRC..... International Federation for the Red Cross
- IDP..... Internally Displaced Person

INGO.....	International Non-Governmental Organization
IOCE.....	International Organization for Cooperation in Evaluation
ISO.....	International Organization for Standardization
M&E.....	Monitoring and Evaluation
MSF.....	Medicine San Frontiers
NATO.....	North Atlantic Treaty Organization
NGO.....	Non-Governmental Organization
NPM.....	New Public Management
NRC.....	Norwegian Refugee Council
OECD.....	Organization of Economic Cooperation and Development
PE.....	Participatory Evaluation
P&O.....	Prosthetic and Orthotic
P-PE.....	Practical-Participatory Evaluation
RBM.....	Results Based Management
RTE.....	Real-Time Evaluation
SGS.....	Société Générale de Surveillance
TOR.....	Terms of Reference
T-PE.....	Transformative-Participatory Evaluation
UFE.....	Utilization-Focused Evaluation
UN.....	United Nations
UNHCR.....	United Nations High Commissioner for Refugees
USAID.....	United States Agency for International Development

Acknowledgements

I offer my thanks to the faculty, staff and fellow students at UBC who have contributed to an intellectually stimulating graduate experience. Particular gratitude is owed to Dr. Sandra Mathison for supporting me at all stages of this project and whose advice, guidance and constructive feedback was crucial in allowing me to complete this project. I am also grateful for the support of Dr. Erin Baines for sitting on the research committee and providing valuable insights into humanitarianism. Thanks are also due to Dr. Wayne Ross for serving as the external examiner for the research project.

I owe a sincere debt of gratitude to the evaluators and INGO staff who so generously gave of their time to talk to me about their experience with evaluation. Their contributions were crucial for understanding the context of the evaluations and spending so much time to talk to me was greatly appreciated.

To my family and friends I am grateful for your support during my studies especially my partner Li who has given unreservedly her love and support.

Chapter 1: Introduction

The desire to help people affected by disasters is a core instinct of human nature that permeates modern society. Pictures of suffering children, refugees fleeing conflict and homes destroyed by floods pepper our televisions screens and elicit pledges of support from governments, foundations and individuals. However the days when humanitarian assistance was seen as an unquestionably positive intervention have long disappeared. Following the widespread problems with the response to the Rwandan genocide, there have since been reports about the misappropriation of funds, aid falling into the hands of soldiers, and the sexual abuse of women and children by aid workers which have all contributed to a debate about the purpose, ethics, and effectiveness of humanitarian assistance. At the heart of this debate is the idea of accountability. Accountability has been a buzzword in the humanitarian relief sector since the mid-1990s, as it has been in international development work and in other areas of public spending. However a key question in this debate is accountability to whom? Who is responsible for accounting for their actions and who are they giving an account to?

Linked to the questions of accountability is the debate about the role of evaluation in this process. In many sectors when evaluation is seen as an accountability tool it is strongly linked to a technocratic, upward form of accountability that has developed from the new public management (NPM) movement of the 1980s and 1990s (Chouinard, 2013). This has in the past also been true in the humanitarian relief context as donors have required verification of performance targets detailed in a project log-frame and evaluation has been seen as the means to provide this. However evaluation has the potential to play a role in accountability to intended beneficiaries, and be used as a means to not only evaluate how accountable an organization has

been in a project but also as a tool to account for the organizations actions to the target community.

This study examines how international non-governmental organizations (INGOs) are using the evaluation of humanitarian action (EHA) as an accountability mechanism. It examines the context and historical evolution of the accountability debate within the humanitarian sector and identifies how evaluation fits into this debate. While there has been an ongoing debate about accountability in the humanitarian sector and a general belief that developing accountability mechanisms to intended beneficiaries is an important task for INGOs, there has been limited research about whether ensuring accountability mechanisms improves program implementation or indeed whether INGOs are making themselves accountable to intended beneficiaries. As such, this study asks whether INGOs are making themselves accountable to intended beneficiaries through the use of evaluation.

Although only one of many accountability mechanisms, the use of evaluation in humanitarian relief has become increasingly widespread since the 1990s (Feinstein & Beck, 2006; Riddell, 2007) and donors often require evaluation in a project plan. However the quality and utility of evaluations has received widespread criticism (Feinstein & Beck, 2006; Riddell, 2007). Evaluation is one of five accountability mechanisms used by NGOs (Ebrahim, 2003) and various evaluation approaches seem readily designed to be used for this purpose. Additionally codes of conduct such as the Humanitarian Accountability Partnership's (HAP) Standard in Accountability and Quality Management identify participation in evaluations and sharing of information through evaluation as being key benchmarks for organizations in ensuring they are accountable to the people they aim to assist.

Despite this progress there is still tension and disagreement within the sector as to whether an ‘accountability’ evaluation can be used as means of providing accountability to target populations. Many publications identify ‘accountability’ as being an external, upward looking evaluation approach not suitable to participatory approaches (Agyemang, Awumbila, Unerman, & O’Dwyer, 2009). This would seem though to take a narrow view of accountability and the potential of EHA. However recent research highlighting new trends such as using participatory evaluation (Feinstein & Beck, 2006), and that some leaders in the field are integrating evaluation into participatory frameworks (Obrecht, Laybourn, Hammer, & Ray, 2012), suggests EHA is being recognized as an effective mechanism to provide accountability to target populations.

Recent debate concerning accountability in humanitarian relief work has focused on two general forms of accountability; upward accountability towards donors and Western countries’ tax payers and downward accountability to affected communities. For the last two decades, dating back to the questions raised about the international community’s response to the Rwandan genocide in the early 1990s, the issue of accountability towards a program’s intended beneficiaries has been hotly debated (Darcy, 2013; Feinstein & Beck, 2006). Humanitarian relief organizations have formed networks addressing these issues and have signed onto codes of conduct and best practice declarations that aim to put accountability to affected communities at the centre of humanitarian relief efforts. Research is emerging supporting the notion that ensuring accountability mechanisms within a program improves its relevance, effectiveness, efficiency and sustainability, leading to greater impact in the long run (Featherstone, 2013), and reports by various networks such as HAP detail the importance of downward accountability.

Research Questions

The research explored the experiences of INGOs in using evaluation as an accountability mechanism, particularly asking whether evaluation is being used in a way that allows affected communities to hold INGOs to account.

The main research question was: How are INGOs operating in humanitarian relief contexts using evaluations as a mechanism to provide and/or assess accountability to affected communities?

Sub questions investigated were:

What particular evaluation approaches are being used with which forms of accountability in humanitarian relief work?

Who benefits from current evaluation practices?

The research questions were framed to identify the reality of evaluation use in field settings. There has been much discussion about the importance of accountability mechanisms in humanitarian relief in recent years and this study aimed to assess whether this discussion is being translated into actual practices related to evaluation. The questions were framed to allow a critical approach looking not only at the practices of evaluation but questions who is benefiting from the current approaches and addresses issues of power built into the system.

The research involved analyzing a sample of evaluation reports obtained from the Active Learning Network for Accountability and Partnership (ALNAP) website. Once an initial analysis of the reports had been completed and data coding into relevant themes, key informant interviews with NGO staff and evaluators connected to the evaluations or the NGOs sampled were conducted. The goal of the interviews was to gain a contextual understanding of the

information provided in the reports and the current understanding of field practitioners to the concept of accountability.

The data collection from evaluation reports and interviews was analyzed within the framework of the accountability debate provided in a literature review of academic and policy papers. The analysis of current evaluation practice was considered within the context of the policy debate amongst accountability focused networks and partnerships such as HAP and ALNAP, and the attention paid to accountability to affected communities among NGO evaluation policies and accountability statements.

The analysis was done within the framework of critical hermeneutics. Hermeneutics is an approach concentrating mainly on the meaning of a text (Myers, 1995). Its focus is on interpretation and involves using the meaning of a text in a dialectic conversation with the parts of the text that make up the whole. Hermeneutics as an approach incorporates the subjective standpoint of the researcher and thus their ordering or interpretation of text is done according to the subject realities they have experienced. Hermeneutics has been criticized for its lack of a critical approach and failing to acknowledge issues of power (Kinsella, 2006). As such critical approaches to hermeneutics have been developed. Critical hermeneutics primary focus is on social reality and intended and unintended consequences (Myers, 1995). This approach allowed a discussion of accountability issues in EHA that considers the subjective understanding of the actors and situated EHA within the historical context the accountability debate. The approach allowed for a critical examination of how issues of power and control affect the intentions of evaluation practice that complemented the interpretive hermeneutic approach.

Humanitarian Relief

As this research focuses on evaluations of humanitarian relief work it is worth considering what humanitarian relief is. Within the larger umbrella of international aid there are two main branches, international development and humanitarian relief. One of the problems with assessing these fields is there is no universally agreed definition on what is included in either definition. In general, humanitarian relief is considered an intervention in response to a man-made or natural disaster focused on “saving lives, alleviating suffering and maintain and protect human dignity” (Global Humanitarian Assistance, n.d.-a). Development has longer term priorities with focus on the alleviation of poverty as opposed to the focus on saving lives in humanitarian assistance. However fitting projects neatly into one box or another is not always possible. Disaster risk reduction projects are considered part of humanitarian relief but often have development elements and the line between when a humanitarian relief project ends and another starts is not always clear.

Humanitarian relief is traditionally given neutrally and impartially to those in most need. This is a particularly core element of the International Federation of the Red Cross's (IFRC) approach and codified in the Red Cross Code of Conduct which formed the basis of the initial attention to accountability to affected communities. This approach has been under considerable strain in the past decade from the actions of various governments. The use of aid as a foreign policy tool by the US and other NATO countries in Afghanistan and Iraq, the militarization of response to disasters such as the Pakistan earthquake and the severe restraints placed on NGOs by certain governments (for example, the Sri Lankan government at the end of the civil war) all place a strain on the notions of impartiality and neutrality.

Most NGOs working on humanitarian assistance, also work in development and their response will merge from one into the other, and sometimes back again if the situation returns to an emergency one. An example of this would be the current response in South Sudan where conflict is breaking out again moving the focus back to the immediate life-saving needs of the population. Is it also the case different NGOs will have different definitions and trigger points for moving between what they consider humanitarian relief and development. Further complicating the picture, funds raised in response to a humanitarian crisis can often be used for development as an agency will not always be able to spend all the funds raised in a large donation appeal. The more a crisis receives exposure on Western news channels, the more likely the crisis is to receive a high level of funding (Weiss, 2013).

In addition to questions about the length of humanitarian response there is also debate about the breadth of humanitarian response. The issue of whether protection, which is the protection of individuals and groups from human rights abuses, is a core element of humanitarian response is part of the complex situation of humanitarian response (Riddell, 2007). This impacts the accountability question as protection concerns can conflict with pledges of neutrality and impartiality. Does accountability go beyond accounting for a project or program to a target population to accounting for how an organization's work impacts the broader human rights situation in a country? The most recent HAP report argued the concern for the wider political impact concern had been lost as recent debate has focused on the relationship between implementing agencies, beneficiaries and the delivery of services (Darcy, 2013). This concern will be discussed further in the literature review section of the study.

Many of the issues related to accountability in the development and humanitarian relief sectors are similar. The notions of power, donor control and mandatory and voluntary accountability are relevant to both sectors, although the historical pace of change and debate is different. The selection criteria for identifying the sample of evaluation reports did not include an analysis of where the intervention occurred along the humanitarian relief/development continuum. If the report had been posted by the INGO on the ALNAP website then it was accepted the organization considered elements of the intervention were linked to humanitarian relief. This approach ensured the selection of reports included a number of interventions for immediate crises such as Action Contre la Faim's (ACFs) response to Hurricane Yolanda, projects focused on longer term crises such as CARE's peace building project in Gaza, and work on post-conflict recovery, which included the Norwegian Refugee Council's (NRC) education response in Northern Uganda.

Research Assumptions

The research questions aim to assess how evaluation is being used and who INGOs are being accountable to through the current system. The research seeks to identify whether current practices support evaluation being used as a mechanism for INGOs to account for their actions to affected communities. One of the assumptions running through most of the codes of conduct and policy level debates amongst accountability advocates is accountability to affected communities is positive and helps to improve programs. In fact, there is limited research suggesting this is the case. Most recently a case study research by Christian Aid, Save the Children and HAP found three projects performed better on the DAC/OECD evaluation criteria when downward accountability mechanisms were in place (Featherstone, 2013), but more research is needed in this area. Despite the limited research into this issue I also assume accountability to affected

communities is a positive approach INGOs should be adopting from a moral perspective as well as practical program improvement perspective. The assumption throughout the research is that holding oneself to account to target communities is an important action for INGOs to take from a democratic standpoint that empowers communities and develops better long-term response programs.

A further assumption I make in this research is when evaluation is used for accountability purposes it can be both participatory and involve lesson learning if the accountability focus is a downward one. This is an issue of debate within the humanitarian community and some evaluators suggest the current accountability focus in the evaluation of humanitarian action precludes downward accountability (Hallam, 2011). Without pre-judging whether the current structures and approaches allow for downward accountability, my standpoint is evaluation can offer this form of accountability when used in a manner that prioritizes the opinions and experiences of the intended stakeholders.

Contribution to the Body of Research

Although there has been much discussion about the need for accountability mechanisms to beneficiaries in the last decade, and the development of several self-regulatory bodies, there has been only limited empirical research as to the practices of INGOs in this area, particularly in the role evaluations play. Generally the work in this area has been a conceptual call for improving accountability mechanisms and improving the quality and participatory nature of evaluations. Analysis of the overall quality of evaluations through, for example, ALNAP's annual meta-evaluation reports has been done. However ALNAP stopped this process after 2008 and replaced with its State of the Humanitarian System report. HAP also publishes annual

reports on the humanitarian system and more specifically focuses on accountability. HAP's annual reports, particularly in 2008, 2009 and 2010 did review evaluations and addressed the issue of accountability, but this was only a small section of the annual reports. The reports included sampling published evaluation reports and analyzing whether the evaluations were interviewing beneficiaries, explicitly considering accountability to local communities, and systematically assessing accountability to local communities (HAP International, 2008, 2009, 2010c). The analysis did not consider whether the evaluations themselves were being used to provide accountability to local communities. This review was not undertaken in the 2011 report. Various academic studies have also addressed accountability and evaluation from a conceptual approach but there has been limited empirical studies on the whether evaluations are being used to enhance downward accountability. As such this study will address a gap in research by adding to the body of knowledge of what is known empirically about the use of evaluation in humanitarian relief, and the possibilities evaluation offers in providing accountability to targeted communities.

Chapter 2: Literature Review

This study examines how INGOs use evaluation as an accountability mechanism in their humanitarian response activities. This review of literature on NGO accountability, how accountability is defined within the humanitarian sector, the impact of neo-liberal policies, the role of power relations in humanitarian responses, and the use of evaluation in the sector helped frame my research.

NGO Accountability

The Oxford English Dictionary defines accountability as “the quality of being accountable; liability to account for and answer for one's conduct, performance of duties, etc. (in modern use often with regard to parliamentary, corporate, or financial liability to the public, shareholders, etc.); responsibility” (OED, 2013).

The literature on NGO accountability includes numerous and varied definitions and is therefore a good starting point as many of the issues affecting humanitarian review are common to the NGO sector as a whole. The various definitions of accountability show the tension of competing demands for accountability, most notably from donors and governments and target communities (Ebrahim, 2003). Accountability can be seen as a formal hierarchical concept where clear lines of reporting are established. Edwards & Hulme (1996) put this as “the means by which individuals and organizations report to a recognized authority (or authorities) and are held responsible for their actions” (p.976). A broader definition sees accountability as both being held responsible by others and taking responsibility for oneself (Cornwall, Lucas, & Pasteur, 2000), introducing the idea that accountability is both regulatory and voluntary. In this definition NGOs report their actions to recognized authorities but also take responsibility

publically themselves for the way they act. These versions are both external forms of accountability. Another approach to accountability is an internal accountability, which in the case of NGOs is accountability to its mission and staff (Ebrahim, 2003; Najam, 1996).

Building on the distinction between formal and voluntary approaches, accountability can also be understood from traditional technical approaches and a stakeholder approach. The traditional approaches fit into formal, contractual approaches to accountability, and are similar to the version of accountability Chouinard (2013) has suggested is linked to New Public Management (NPM). The stakeholder approach sees accountability as a right, and this right is held by anyone affected by the program or NGO's policies (Lloyd, 2005). This approach fits within a partnership and offers a more inclusive definition of accountability.

Accountability is not just about being responsible, but about being seen to be responsible (Wenar, 2006). In this sense, accountability has a visibility requirement. It is not enough to achieve project goals, an organization has to demonstrate they are effective. Wenar (2006) claims accountability is directional and the group or individual holding a person or organization to account have certain powers. These include the power to judge, the power to set certain standards and the power to sanction if standards are not met. These powers are problematic for downward accountability since intended beneficiaries rarely have formal powers to sanction NGOs for their performance (Darcy, 2013; Lloyd, 2005).

These definitions focus on the responsibility to account for one's actions. Accountability can also be understood to be the responsibility to act in the first place. This is particularly relevant in humanitarian relief where the lack of action has often been a criticism. In this understanding of accountability, organizations must not only account for their actions, but also

be accountable for acting when they have a duty to do so. They therefore can be called to account for why they did not act in a particular situation (Gray, Owen, & Adams, 1996).

What this initial introduction shows is the complexity of accountability. Different approaches and understandings complicate attempts to define accountability for NGOs. An INGO may have to balance different stakeholders' understanding of accountability, and competing demands from various stakeholders can be severe and time consuming (Ebrahim, 2003). As well as intended beneficiaries and donors, host governments, taxpayers in donor countries, INGO staff, other stakeholders and communities in recipient countries, partner organizations and other members of certification/accountability groups also make accountability demands of INGOs (Jordan, 2009; Lloyd, 2005). Another way of looking at stakeholders is who is asking the accountability questions and Jordan (2009) ranks stakeholders based on this factor. Donors are ranked at the top followed by other NGOs and host governments; beneficiaries ask the fewest accountability questions.

The most common directional split in accountability is upward and downward accountability. In humanitarian relief, upward accountability is being accountable to donors and often comes with a contractual obligation. It can also refer to accountability to host governments, the media and taxpayers in donor countries (Cosgrove, 2007). Downward accountability is accountability to affected communities and it often considered voluntary or a moral accountability (Ebrahim, 2003; Lenn, 2006). Horizontal accountability is accountability to other organizations working in humanitarian relief, and inward accountability is accountability to staff and members for upholding values and missions statements (Lenn, 2006).

The Evolution of Accountability Within the Humanitarian Sector

The recognition of the need for greater accountability to affected communities dates back to the 1990s. The catalyst for the movement towards affected community accountability was the Rwandan genocide and the response to the emergency by the humanitarian community (Borton, 2004; Darcy, 2013; Feinstein & Beck, 2006).

Disaster Responses and Accountability

The responses to the 1994 Rwandan genocide led to the largest and most expensive evaluation of a collective humanitarian response action ever undertaken (Borton, 2004). The evaluation component that had the most impact was the section on accountability and Borton (2004) argued at least three accountability initiatives, the Sphere Standards, the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) and HAP stemmed directly or were strongly influenced by the joint evaluation. These three initiatives, along with the Red Cross Code of Conduct and the launch of People In Aid, are probably the most influential within the humanitarian sector (Everett & Friesen, 2010). The evaluation also addressed the issue of the need for policy debate on preventing genocide, strengthening human rights protection in Rwanda, and developing an early warning system for conflict in the Great Lakes Area. There is less evidence these recommendations had an impact on policy and practice (Borton, 2004; Walker & Maxwell, 2009).

The certification process launched by HAP and People in Aid are supported by the largest INGOs. However, numerous other initiatives have also been launched. Everett and Friesen have counted at least 24 different accountability mechanisms and introduce the idea of “multiple accountability syndrome” (Everett & Friesen, 2010, p. 474), whilst Alexander (2013) found even more; 145 codes, initiatives and standards. The approaches to accountability vary in these codes

and initiatives. Some, such as the Sphere Standards, focus on technical standards with measurable indicators, whereas others such as the Red Cross code are a statement of values and ideals. People in Aid is focused on accountability to staff members, and the accountability comes from having qualified staff to implement a project. The HAP initiative is firmly focused on downward accountability. As the issue of accountability became more prominent in humanitarian relief, the development of accountability systems became an industry within an industry with initiatives competing for the attention of international organizations. The initiatives utilize funds that might otherwise be used to fund program delivery both from running the initiatives themselves and the resources organizations invest to ensure they are compliant with one code or another.

The second major event that fuelled further introspection within the humanitarian community was the 2002 United Nations High Commissioner for Refugees (UNHCR)/Save the Children report on the sexual exploitation of refugees by aid workers in West Africa (HAP, 2006). This report brought the issues of protection and human rights firmly into focus for the accountability movement. This framed accountability not just as accounting for how a project had met its targets, but also as taking responsibility for the behaviour of individuals such as staff members who are in a position of power.

Two 2005 events were critical in shaping accountability in the humanitarian sector. The first was the response to the 2004 Asian Tsunami, which was the largest scale natural disaster to be highlighted in the modern age of communication and twenty-four hour news. NGOs and UN agencies were inundated with funds and in the rush to respond, the needs and views of the target populations were often ignored (Riddell, 2007). The Asian Tsunami did not have the same impact as the Rwandan genocide in producing codes of conduct and accountability initiatives,

but encouraged reflection on the challenges of accountability still facing the sector over a decade after the Rwandan response. This event highlighted the challenges of an increasing competitive and crowded field with new, and often inexperienced, organizations responding to disasters. A joint evaluation of the response found the dominant accountability focus was upward to donors (Cosgrove, 2007), in spite of the initiatives launched in the previous decade.

2005 also saw the UN make high level commitments to the issue of accountability, mirroring the commitments made by many NGOs in previous years (HAP, 2006). The UN is a major player in the humanitarian system and particular agencies such as UNHCR are especially important in the response to armed conflict. Coordination for response is often required to go through the UN by host governments, and the UN chairs various clusters (such as water and sanitation or health). Despite this important role, which gives the UN considerable influence over the lives of victims of disasters, the acknowledgement in 2005 of the importance of accountability to affected communities were the first made by senior UN figures (HAP International, 2006). The UN is a large bureaucracy where the rate of change can be slow and the highlighting of accountability to affected communities by senior UN officials can be seen as an indicator the accountability movement was becoming increasingly recognized throughout the international community.

Other major disasters have occurred since 2005, most notably the 2010 Haitian earthquake, Hurricane Haiyan in the Philippines, the internally displaced person (IDP) crisis in Sri Lanka at the end of the civil war, and the current civil war in Syria. These events have or will have large scale evaluations of the responses. They will also continue the accountability debate, although none have as yet produced the same level of influence as the reflections and evaluations of the response to the Rwandan genocide. Concerns about the response to armed conflict has also

raised questions of accountability to the wider strategic developments, specifically whether humanitarian aid can prolong wars by changing the resource and power dynamics and supplying aid indirectly to armed groups (Okumu, 2003).

Tracing Accountability Through the Humanitarian Accountability Reports

Tracing the major themes and issues that are addressed in the Humanitarian Accountability Report produced by HAP illustrates how the debate over humanitarian accountability has changed more recently. Between 2005 and 2011 HAP's annual report, written by an independent consultant, gave an overview of humanitarian accountability for the year. There were also brief reports from each member on their progress towards accountability. HAP did not publish a report in 2012, but did publish a detailed review in 2013 to mark HAP's 10 years of operation. This report reviewed progress over the entire ten year period and included reports by several independent consultants, rather than just one. The 2005 HAP report defines humanitarian accountability as "giving beneficiaries a proper say in humanitarian aid" (HAP International, 2005). As such the reports look at the progress in downward accountability.

As one of the leading forces driving the discussion of accountability within the humanitarian world, reviewing the annual reports illustrates how accountability issues have changed recently. The first report in 2005 focused heavily on the accountability deficit, or the imbalance in accountability to beneficiaries and accountability to donors. Darcy (2013) defines the deficit as "the lack of attention given by humanitarian aid organisations to aid recipients as compared to 'official' stakeholders-donor governments and others." (p.5) The report argues the response to the Asian Tsunami had ensured the accountability deficit problem, which was already recognized by the humanitarian community, was brought to the forefront of discussions. The report also noted the limited impact of the accountability initiatives in the field, and painted

a bleak picture of the state of accountability within the sector. This report clearly demonstrated that in 2005 at least, the humanitarian sector practice of accountability was still dominated by an upward model where donors held the power and populations affected by disasters struggled to make their voice heard. It also appears this view was widely acknowledged within the humanitarian sector. In a survey conducted by HAP, 59% of humanitarian professionals identified organizations were regularly making themselves accountable to donors, compared to only 8% identifying the same for beneficiaries (Alexander, 2013).

Whilst the accountability deficit has remained an issue in the HAP annual reports, the main themes have evolved year by year. The 2006 report highlights the lack of enforcement mechanisms in accountability initiatives and the concern negative press reports on aid give NGOs' communication teams leverage over accountability officers as NGOs seek to control the publication of potentially negative information (HAP International, 2006). By 2007, the report moves the debate away from a general discussion about the need for more accountability and to a review of how effective the existing accountability mechanisms are at addressing power imbalances in the humanitarian world (HAP International, 2007).

The 2008 report introduces themes that would be key in future years. In particular, the report focuses on the multiple accountability initiatives that exist and their impact on accountability in general, and linked to this, the national level certification schemes in countries such as Cambodia and Pakistan. For the first time, the 2008 report included a brief assessment of evaluations placed on the ALNAP website for evidence of accountability (HAP International, 2008). Attention to the issues of multiple initiatives and the standards of evaluation continued up to the 2011 reports (HAP International, 2009, 2010c, 2011). A major issue raised in 2010 was the

use of information communication technology (ICT) to improve accountability through the flow of information to intended beneficiaries (HAP International, 2010c).

The most recent report in 2013 was a reflection on the progress of the accountability movement in the last 10 years and shifts the debate towards a more wholistic view of accountability than seen in the previous reports. This longer report includes two reflective essays rather than one. In the first, Darcy (2013) argues the debate about humanitarian agencies needing to develop accountability mechanisms has led to a narrow definition of accountability, one that focuses particularly on the responsibilities of individual agencies to implement their projects in an accountable manner. He suggests this limits a strategic approach of accountability that would see aid agencies take responsibility for the impact their actions have in a wider political and societal context, and argues the current approach weakens citizens' abilities to hold their governments to account. The second essay is similar to previous HAP reports but it reviews the evolution of accountability over 10 years, rather than reviewing just the important events in the previous year as the other HAP reports had done.

Summary

The accountability debate has become more nuanced in recent years with discussion moving from a general idea of the accountability deficit to more strategic discussions of humanitarian organizations actions. This may be as a result of greater attention being paid to accountability to affected communities which has seen perceptions of the accountability deficit shrink. The 2005 HAP survey of the perceptions of head office and field staff found only 8% of respondents considered accountability to intended beneficiaries to be high compared to 59% who thought accountability to donors was high. By 2013 the survey results had changed to 51% and 80% respectively (Alexander, 2013). Both upward and downward accountability are perceived to

be increasingly important. The survey relies on self-reported perceptions and may reflect policy level discussions more than actual field realities. The evidence from project sites is less convincing where intended beneficiary consultation and involvement in projects is still limited (ALNAP, 2012).

Accountability Organizations and Codes of Conduct

Horizontal accountability, or initiatives signaling a commitment to accountability to partner organizations, can be seen in codes of conduct and accountability initiatives that have spread through the sector since the early 1990s. The initiatives may change practice improving downward and upward accountability, and also be a signaling mechanism to donors about a commitment to accountability. Five of the most influential initiatives in the humanitarian community are the Red Cross Code of Conduct, ALNAP, HAP International, People in Aid and the Sphere Standards (Everett & Friesen, 2010).

The Red Cross Code of Conduct

The Red Cross Code of Conduct is the only initiative whose inception dates back to before the Rwandan genocide. In progress since 1991, it was presented to the humanitarian community in 1995 (Walker, 2005) and formulates principles for humanitarian action in 10 articles. Leading with traditional humanitarian principles such as neutrality and independence, it also includes newer principles such as participation, accountability and long-term risk reduction (Hilhorst, 2005). The code is a voluntary code without an enforcing mechanism and has been signed by 522 organizations (International Red Cross and Red Crescent, 2014).

The Sphere Standards

The Sphere Project's Humanitarian Charter and Minimum Standards in Humanitarian Response was first disseminated to agencies in 1999 and have since been revised (HAP

International, 2005). The Sphere Standards set minimum standards of delivery for aid in five areas; water supply and sanitation, nutrition, food aid, shelter and site planning, and health services. The standards are technical indicators, for example, an agency responsible for water in a refugee camp should ensure a minimum of 15 litres per person per day are available (The Sphere Project, 2011).

ALNAP

ALNAP's main focus is facilitating learning and knowledge sharing in order to improve accountability and humanitarian performance. ALNAP is a membership based organization which unlike the Red Cross Code of Conduct includes donors and UN agencies. One of ALNAP's mechanisms for encouraging learning is the storing of evaluation reports on a publically accessible website. It also conducted meta-evaluations of the evaluations on its database before switching in 2012 to a 'State of the Humanitarian System' report.

People in Aid

The People in Aid and HAP initiatives have a regulatory element to them and thus differ from the other standards. People in Aid focuses on improving human resource management of humanitarian and development NGOs and emphasizes both accountability of NGO staff and indirectly downward accountability by ensuring NGOs have staff capable to implementing programs for intended beneficiaries. People in Aid has a code of good practice and a certification process for members.

The Humanitarian Accountability Partnership

HAP's forerunner was the Humanitarian Ombudsman Project, which relied heavily on compliance with the Red Cross Code of Conduct and the Sphere Standards. The approach foresaw a system-wide ombudsman system acting as a regulator of accountability standards. This

was met with resistance and was perceived as a heavy-handed, wrong approach and closed after two years and HAP replaced it in 2003 (Hilhorst, 2005). The ombudsman system was considered infeasible to implement given the fractured nature of the humanitarian system and the number of different players. International organizations also raised questions about the legitimacy of the system. HAP's focus is almost exclusively on downward accountability and like People in Aid has a certification process. HAP certifies a NGO is compliant with its 2010 Standard in Accountability benchmarks in six areas: establishing and delivering on commitments, staff competency, sharing information, participation, handling complaints, and learning and continual improvement (HAP International, 2010b).

There are many other standards and certification processes. A review of recent HAP reports illustrates the development in both global and national level schemes. The 2013 report lists 42 initiatives, 41 standards and guidelines and 64 codes of conduct (Alexander, 2013). The proliferation of these schemes demonstrates accountability has become important and a standard definition of accountability in the humanitarian sector is still elusive. A major concern within the sector is the proliferation of accountability mechanisms will either lead NGOs to spend too much time fulfilling the requirements of multiple schemes, or they will pick one certification process that focuses on one form of accountability at the expense of others (HAP International, 2008).

Power and Accountability

The humanitarian sector clothes itself in the language of independence, impartiality and neutrality. However, all humanitarian work seeks to change the current situation on the ground. Both man-made and natural disasters create situations that individuals, organizations and states within the humanitarian sector identify as needing help and support to change the situation. A traditional humanitarian would argue the change is simply to alleviate suffering but this view

ignores humanitarian relief's impact on policy, services and resources. Humanitarian intervention has the potential to distribute resources, change national and international policies, affect the relationship of a community to the state and impact status and authority locally and nationally (Kennedy, 2004). Power is therefore a central element in the debate about accountability in the humanitarian sector, which has become increasingly important as the neutrality and independence of humanitarian actions is threatened by the policies of donor governments.

There has been an increased movement to align the use of humanitarian aid with national interest. The idea of humanitarian intervention developed after the end of the Cold War and has been criticized for using humanitarian motives as a cover for furthering national interest (Gibbs, 2009). The Afghanistan and Iraq conflicts and the 'War on Terror' in this century have further fuelled the belief that aid is a tool for pursuing national interest. Advertising aid as a 'hearts and minds' battle has become dominant, threatening the independence and often safety of aid workers (Burns, 2010). Illustrating that aid is becoming increasingly dependent on foreign policy, the Canadian Government announced in 2013 it would absorb the Canadian International Development Agency (CIDA) into its Foreign Affairs Department. This idea was copied in September 2013 by the Australian Government (Santamaria, 2013).

The alignment of aid with national interests harms both the neutrality and need-based principles in humanitarian relief. INGOs that rely on government funding may have their neutrality questioned and may be forced to make the choice between policy or advocacy decisions that go against basic humanitarian principles or losing funding, and perhaps in some cases in Canada, even losing their charitable status (Barry-Shaw & Jay, 2012). A needs based approach is challenged by the alignment of aid with national interests. Billions of dollars have

been poured into humanitarian relief and development in Iraq and Afghanistan while at the same time conflicts in places such as the Democratic Republic of the Congo have been largely ignored (Hilhorst, 2005). As such, aid is not apportioned on a needs basis but an interest basis.

The consolidated appeals process (CAP) is an attempt by the UN to coordinate appeals for funding from its agencies and INGOs to donor governments. Comparing the percentage of requested funds for identified needs to actual donations is a good way of highlighting how national interest can outweigh humanitarian need. In 2011, the highest ranked countries of percentage of needs met were Afghanistan, Libya, Sudan, Sri Lanka, and Haiti. The lowest percentages were Zimbabwe, Djibouti and the region of West Africa (Weiss, 2013). Afghanistan and Libya have both been the scenes of NATO interventions, and Sudan, an oil rich country, has received the attention of powerful celebrities such as George Clooney. Sri Lanka is a popular tourist destination whose civil war was featured heavily in Western news reports, and Haiti's earthquake also received blanket news coverage and is geographically close to the U.S. Zimbabwe and Djibouti, and West Africa in general receive limited news coverage and are considered less important to the national interest of Western governments.

With this in mind, one needs to ask how power affects the accountability movement. HAP International's standard in accountability and quality management defines accountability as "the means through which power is used responsibly" (2010b, p. 13). This definition is included in the opening section of the 2010 HAP Standards in Accountability and Quality Management. This section also identifies accountability as a right for "anyone affected by the use of authority" (2010b, p. 13). Accountability is a means for holding those with power responsible for their actions: "Broadly speaking, accountability refers to the process of holding actors responsible for their actions. More specifically, it is the concept that individuals, agencies and organizations

(public, private and civil society) are held responsible for executing their powers according to a certain standard (whether set mutually or not)." (Tisne, 2010, p. 2)

Whilst accountability can be seen as a means of holding those in power responsible for their actions, it can also be used as a means of consolidating power by those who hold it. In complex hierarchical systems, power holders delegate authority to others to implement their desired policies. In this type of system, accountability becomes the means by which power holders ensure those they delegate responsibility to, justify their actions (Mathison & Ross, 2002).

Accountability can be used both to hold those in power responsible for their actions (or lack of actions) and also by those in power to hold those they delegate authority to responsible for implementing their policies. The practice of accountability has two elements to it; answerability and enforceability. Answerability involves those who implement programs and policies giving an account and justification of their actions. Enforceability involves the possibility of sanctions or repercussions for a failure to account adequately for actions or failings (Goetz & Jenkins, 2005). To understand the system of accountability and the context with which EHA operates it is important to understand the sources of power in the humanitarian system, who has the power to demand answerability and deliver enforceability, and how the relative importance of each affects different actions in the system.

Influence is a key part of the power dynamic. Traditionally the stakeholders who can influence decisions that an INGO make are the recipients of upward accountability. There is though a large group of stakeholders who are affected by the decisions of the INGO who do not have influence (Blagescu, Casas, & Lloyd, 2005). The most relevant example in this case is the community the INGO is working in. The principle behind humanitarian accountability is

reducing the power imbalances between groups with influence and those without. An accountability approach therefore that was truly attempting to reduce power imbalances would have consider how all stakeholders could have influence over decisions that affect them (Blagescu et al., 2005).

Donor Governments

Humanitarian relief is a business and aid can have major impacts on the communities and countries where it is distributed (Kennedy, 2004; Weiss, 2013). INGOs and local NGOs compete for that business. The competition has increased dramatically in the last 20 years with global spending increasing almost four-fold between 1980 and 2005 (Riddell, 2007). Although private donations, and increasingly corporate spending, make a significant contribution to humanitarian relief spending, the majority of spending comes from donor governments. There are other sources of power within the humanitarian system, but controlling access to funding and resources is the dominant form of power. Maintaining the necessary levels of funding for an organizations' programs can be a constant struggle (Barnett, 2011), and this hands significant power to a donor government in holding a NGO to account. The movement to bilateral giving, which means more funds are given directly to NGOs rather than multilaterally through the UN, has also increased the strength of the donors' position (Riddell, 2007; Weiss, 2013). Donor governments are therefore in a position to both demand answerability and administer sanctions or repercussions. Donors can refuse to fund future projects, suspend funding from NGOs or even withhold funding for a national government if they feel their demands for accountability have not been met.

Donor governments also hold power through the traditional inter-governmental structures of modern diplomacy such as the UN and the Bretton Woods institutions. The largest donors of humanitarian aid are some of strongest states—economically, politically and militarily. This

gives them power in setting and implementing policies that impact the humanitarian sector (Riddell, 2007).

Donor governments face various challenges for accountability. A key location for accountability demands comes from a domestic audience. Taxpayers, special interest groups and the media can demand accountability in the humanitarian sector (HAP International, 2005). In this context, these groups can both demand a government accounts for its actions and that a government holds NGOs or the aid world in general to account. This in itself can create a dynamic where accountability is actually reduced rather increased. Aid agencies have a vested interest in portraying themselves in the best possible light. This is particularly the case in the current financial crisis where there is considerable pressure on governments to cut foreign aid rather than make domestic budgetary cuts. In a world where the average American thinks foreign aid is 10-25% of the U.S. federal budget, when it is in fact 1% (Melamed, 2014), the incentive to avoid negative news stories is strong. As such INGOs can be reluctant to publish negative evaluations (HAP International, 2006; Riddell, 2007) which reduces rather than increases opportunities for accountability.

Other attempts to hold donor governments accountable for their actions and promises have also been made recently. The Good Humanitarian Donorship (GHD) meeting in 2003 led to the development of Principles of Good Practice which included mechanisms to improve donor accountability (Alexander, 2013). Initiatives such as the Humanitarian Response Index (HRI) by Development Assistance Research Associates (DARA), the International Aid Transparency Initiative (IATI), and Publish What You Fund, have highlighted this approach to accountability of donors. HRI ranks donors performance against the GHD principles. The introduction of the HRI was met with resistance by donor governments, particularly those that scored badly on

performance (HAP International, 2008). Unlike initiatives such as HAP and ALNAP, HRI is funded without government donor assistance and so has the potential to maintain independence. HRI reports to date have raised questions about the lack of transparency of decisions on fund allocation and observed funding was not based on humanitarian principles (Alexander, 2013).

Non-Western donor governments are becoming increasingly active in humanitarian relief. Obtaining data on the amount of funds given is difficult as these new donors do not necessarily follow the same accounting policies as traditional donors. However countries such as China, Saudi Arabia, Brazil, Turkey, and India have all made increased contributions to humanitarian causes recently (Binder & Meier, 2011). Motivations for giving vary, but as with Western donors, national interest and global reputation are important. Non-Western donors are more likely to direct money through recipient governments' treasuries. This comes from the different approaches to state sovereignty of the countries involved. Traditional donors have more and more followed the 'Responsibility to Protect' doctrine since the 1990s that argues state sovereignty should not be a barrier when states fail in their obligations to protect their citizens. This has increased the alignment of military and humanitarian policies in these countries and been the justification for military engagements (Weiss, 2013). This doctrine is rejected by many of the newer donors and has led to both a high proportion of funds being directed through state treasuries and a limited funding of protection/human rights programs (Binder & Meier, 2011). The increased presence of non-Western donors therefore has the potential to change the power balance in the humanitarian world. Western donors may see a reduction in their influence at the expense of other donors, and recipient governments may have greater control compared to INGOs and UN agencies. How this affects the ability of beneficiaries to hold various actors to account is not yet clear.

Host Governments

Humanitarian aid differs from international development aid because funds are funneled more to INGOs and UN bodies rather than national governments (Darcy, 2013; Riddell, 2007). The Paris Declaration on Aid Effectiveness committed donor governments to build the capacity of recipient governments by directing as much as aid as possible through the recipient government's treasury. However, as a sub-sector of international aid, humanitarian relief was not particularly affected by this as it is widely delivered through INGOs and UN agencies (Darcy, 2013).

Within the overall structure of the aid system, the Paris Declaration intended for mutual accountability to be undertaken between donor and recipient countries. The concept of mutual accountability rests on the idea that two partners will hold each other to account, and the accountability mechanisms set up in each country will complement mechanisms in the other country reinforcing accountability and transparency. This arm of the Paris Declaration has been the least successfully implemented (Jones & Picanyol, 2011). Part of the reason for this is there is an asymmetrical power relationships between donors and recipient countries, in that “donors typically can withhold aid, while aid recipients have few means of exerting influence over donors” (Jones & Picanyol, 2011, p. 6). Recipient countries therefore have limited powers over donors on international aid in general, and this is even more limited in humanitarian relief where the funds are less likely to be channeled through their government’s treasury. As noted, the entrance of non-Western donors may alter the power imbalance, but the effects of this are not yet clear.

The funneling of funds through INGOs and UN agencies impacts the possibility of citizens of a recipient country to hold their government to account. Article 6 of the Red Cross

Code of Conduct emphasizes using local capacity to respond, but this article has been identified as the most ignored article (Hilhorst, 2005). Darcy (2013) highlights how this changes the accountability dynamic. Under the international system, the primary responsibility for the protection and support of a country's citizens rests with the country's government. As such, a government has the responsibility to ensure access to food, shelter and health care in an emergency. A government that failed to do so could be held to account by its citizens in whatever ways the political structure in the country allowed. However, under the current humanitarian system, this responsibility is often assumed by UN agencies and INGOs. In Darcy's argument this weakens the ability of citizens to hold their own governments to account. Instead local citizens look to UN agencies and INGOs to account for their actions, which these agencies do out of a moral sense of responsibility not a contractual obligation. Darcy sees this as a weakness of the accountability movement. Organizations consider their accountability responsibilities to be limited to the delivery of the programs they are contracted to implement. By failing to consider the wider political context, they are failing to consider their programs are displacing the responsibility of local authorities to provide services and make decisions on resource allocation. This then removes power from citizens of the country to demand basic services from their government.

The power relationships between host governments and INGOs are complex and quite context specific. In humanitarian relief, INGOs often hold access to the funds to provide services and the expertise to do so. They may also have access to policy makers both in donor governments and so be able to advocate for prioritization of particular funding streams, and in host governments and so be able to affect national policy.

What INGOs do not have without the compliance of host governments is access to the people they wish to serve. Strong, authoritarian governments may limit access to beneficiary populations. This may be just a blanket refusal to allow meaningful access as in the case of Burma after Cyclone Nargis in 2008 (Human Rights Watch, 2010). In other cases, access may come with conditions that often leave INGOs with a moral dilemma of compromising to reach a population or sticking strongly to principles which results in them losing access. For example, Sri Lanka forced Tamils displaced at the end of the civil war to live in IDP camps where permission to leave was hard to obtain and services and conditions were poor. INGOs and UN agencies were granted very limited access (Human Rights Watch, 2009). This left INGOs to either operate in the camps and thus semi-legitimizing the Sri Lankan Government's policies, or abandon the populations in the camps.

This is an example of where the expectation of INGOs and the UN to provide services, can reduce a populations' opportunities to hold their government to account. Host governments are able to use their position to blame INGOs for failures of delivery whilst making it difficult for them to do their job. In Sri Lanka the power of Tamils to hold the Sri Lankan Government to account for failing to provide basic services was already very limited in the aftermath of the civil war. By relying on international organizations to provide basic necessities, and blaming them for failures to deliver allowed the government to avoid accountability questions about their own action.

INGOs

INGOs themselves hold considerable power. Their policies and field operations have a considerable impact on the lives of the communities they work with. They hold the key to access to services, defining who are the most at risk members of the community and thus who is

prioritized for attention. INGOs are often major service providers in countries affected by disaster, taking on roles traditionally provided by government such as health-care and education provision. The provision of services and the financial injection they make to a community is not the only source of power for INGOs. They are also able to influence national and international policy. INGOs are often heavily involved in advocacy campaigns to hold governments to account for their promises or lack of action. The International Campaign to Ban Landmines (ICBL) is one example of INGOs using their position as field practitioners and leveraging their moral sway to push for the formulation of a major international treaty (Cameron, Tomlin, & Lawson, 1998).

The ICBL gives an example of INGOs using power they have obtained from positioning themselves globally as moral guardians of the humanitarian world. Wrapping themselves in the traditional concepts of humanitarianism, INGOs try to gain the moral high ground by representing the voice of the marginalized, poor and victims of natural disasters and conflict as a counterpoint to the interest laden approach of donors, national governments and belligerents in conflict. This view of INGOs has been questioned much more forcibly since the 1990s fuelled by concerns of value for money, behaviour of aid workers, and questions about whether aid contributes to war and violence by funding belligerents (Barnett, 2011; Nutt, 2011; Weiss, 2013). However INGOs still retain certain levels of moral legitimacy which gives them power and influence in decision-making. The strength of this power will vary from organization to organization and is often dependent upon the context of the situation. Large organizations that are renowned for taking strong public positions often have greater flexibility than smaller organizations to resist the requirements of large donors and host governments. However this power should not be overstated. In an increasingly competitive market for funding, donor

governments still represent the largest funding opportunities thus reducing the potential of INGOs to push back against donor power.

Populations Affected By Crisis

What power do beneficiaries of humanitarian relief programs and the communities they live in have to hold organizations and governments to account? Under the traditional concepts of accountability they have very little power. Funding for projects ultimately comes from taxpayers of another country, whether it be government or private funding. Affected communities should be able to hold their own governments to account and influence the actions of INGOs this way. This often does not happen for a number of reasons, including a government's failure to provide services in the first place or a government contributing to the humanitarian crisis either through war or policies exacerbating poverty, starvation or disease. As discussed, the presence of international organizations may even weaken a citizen's ability to hold his/her government to account by displacing government services.

Therefore without other ways of acquiring power, affected communities' ability to hold governments and international agencies to account is severely limited. One of the purposes behind the accountability initiatives that have spread through the humanitarian relief and international development sector is to provide these means. This approach comes from the theory of social accountability and the notion that democracy, empowerment and rights based approaches will enhance the ability of aid recipients to ensure they can access needed services (McGee & Gaventa, 2011). The international development field is more sophisticated in these forms of accountability mechanisms. Citizenship voice initiatives, which focus on increasing the opportunity and capacity for people to have their opinions heard and demand accountability from those in power (ODI, 2007), work on the assumption that the ability of a citizen to voice their

views freely improves governance and increases the demand for transparency and accountability (Rocha Menocal & Sharma, 2009b). A review of published reports and evaluations shows this is an approach applied in development but not humanitarian relief settings, and its focus on poverty rather than life saving and relief of suffering emphasizes this. However, the humanitarian relief and development programs often operate in parallel contexts. Development work can therefore impact the response to a crisis. Citizenship voice programs operating in countries prone to disaster have the potential to impact the response to a crisis by nature of having empowered a population to engage their government and thus hold them to account for their response to a crisis.

Citizenship voice projects work to empower communities to demand accountability and rights from their governments. They promote an environment where the demands for accountability will be citizen-led and seen as an integral right (Rocha Menocal & Sharma, 2009a). Efforts in the humanitarian sector have focused on developing structures which give beneficiaries voice to impact the work of INGOs. Identifying and implementing feedback mechanisms such as beneficiary input into programs and complaints procedures is given considerable attention in the humanitarian accountability literature. Research using cases studies on the process and impact of feedback mechanisms has been undertaken recently (Featherstone, 2013; Jean & Bonino, 2013). Implemented effectively the evidence suggests they build participation and improve INGOs' accountability. However they are implemented voluntarily and rely on a moral approach to accountability by INGOs rather than a statutory or contractual requirement.

Participation does not necessarily provide accountability, and accountability is possible without participation. However participation helps strengthen accountability approaches and

ensure greater influence of the affected community. Effective participation allows INGOs to understand the affected community's interests and ensure needs are responded to (Blagescu et al., 2005; Hilhorst, 2002). Accountability to affected communities is enhanced by participatory approaches that ensure communities can influence decision making (Blagescu et al., 2005). This approach reduces the power imbalances which exist within the humanitarian sector. One of the criticisms of INGOs though is participation is often token in nature and does not involve anything more than extracting data at community meetings that rarely translates into impact into decisions (Alexander, 2013; Anderson, Brown, & Jean, 2012).

Donors have become increasingly interested in participatory approaches and beneficiary involvement (Alexander, 2013), and this gives the potential for these mechanisms to become compulsory rather than voluntary. Donors already require organizations to detail how they developed their projects with participatory input and then expect them to report on this throughout the project. How much this has altered the behavior of INGOs and whether donors have moved from policy statements to actually enforcing these ideas in the field is debatable (Alexander, 2013). However should the trajectory of change continue there is the possibility for donors to delegate power to beneficiaries rather than INGOs as is currently the case.

Citizens of Donor Countries

Grassroots movements in the traditional donor countries are another means where the public can hold governments, corporations and international organizations to account for actions or lack of actions. Within the humanitarian sector, the most widely recognized public movement to force greater intervention in a humanitarian crisis was Band Aid; the series of events organized by Bob Geldof and other musicians to raise awareness of the famine in Ethiopia. Serious questions have since been raised about the intervention in Ethiopia and the effect of the

public pressure caused by Band Aid. However there is little doubt the public movement to intervene had a huge impact on the available funds and the willingness of governments and the humanitarian community to respond (Walker & Maxwell, 2009). Band Aid is an example of holding the humanitarian community accountable for their responsibility to act rather than accountable for how they implement programs.

The failure to act is probably the way citizens of donor countries are most likely to hold INGOs to account. International organizations can also be held to account by citizens of donor countries for performance and financial probity. Media reports are generally the most normal form of this. The media will focus on a particular issue or cause that creates a swell of public opinion on a particular issue. Other means include guides to giving such as Charity Navigator which contain details of different non-profits activities and financial performance. Both media reports and websites such as Charity Navigator have subjective ideas of what constitutes good performance and may not be grounded in field realities. This may however reduce accountability as INGOs focus on ensuring negative information does not become available to the press rather than lesson learning through accountability (HAP International, 2005).

Private Donors

Private donations form a significant proportion of humanitarian relief aid. Calculating totals for aid in general, and how much of this is private donations, is difficult but one estimate puts the amount at 30% of all humanitarian aid managed by delivery organizations (i.e. organizations working on project implementation in a crisis zone) (Global Humanitarian Assistance, n.d.-b). The amount of private donations a particular NGO will receive is dependent on the individual organization. MSF has tried to carve out an identity particularly focused on

independence and its private fundraising is particularly impressive. Other organizations, such as the NRC are more comfortable being mainly funded through governments (Weiss, 2013).

The source of private donations is diverse and includes general and crisis specific fund-raising campaigns with the general public, donations from churches and other religious organizations, and private corporations. The diverse nature of funding and the way it is used often lessens the power of the giver to hold an INGO to account for their actions, and government donors have created a system to compound this problem. Government donors give very limited, if any proportion at all to support head offices. However an INGO will often need a head office to perform functions such as human resources, donor management, and policy development. INGOs therefore use private donations to fund their head offices and prefer donations and bequests are not earmarked for a particular location or crisis (Nutt, 2011). Even with donations given for particular causes, it can be hard to hold an organization to account. Individuals and often also religious and non-religious foundations have less access to on-the-ground knowledge than donor governments and therefore more reliant on feedback from the INGO and media stories to assess the impact of their donations.

The HAP surveys on perceptions of accountability affirm the argument government donors are considered more of a priority for accountability than private donors. In the 2011 survey, 81% of respondents considered the level of accountability given to government donors to be high, compared to 66% for private donors. These percentages were 59% and 23% respectively in 2005, suggesting the significance of private donors to humanitarian organizations has grown. The figures in 2011 for beneficiaries, host governments, and the general public was 41%, 41% and 29%, suggesting despite the various movements for accountability, funding was still the leading force behind accountability in 2011. The survey does not make clear whether the

respondents are perceiving this as all private donations, or large-scale donations from foundations or corporations.

If the role of private foundations is becoming increasingly important in humanitarian relief this does have the potential to change power dynamics in the sector but this may just shift the power from Western governments, to western corporations and foundations controlled by religious organizations or rich individuals. A number of corporations such as oil, tobacco, logistics and financial companies have developed corporate giving programs and formed partnerships with UN agencies, the Red Cross or INGOs (Van Wassenhove, Tomanisi, & Stapleton, 2008). There has been limited research on how this impacts the accountability debate and who benefits from this development. It also opens the wider debate about concerning the power of corporations in donor countries that is beyond the scope of this discussion.

Religious organizations are also heavily involved in humanitarian relief, and a number of INGOs are faith-based. World Vision for example is an evangelical Christian organization that is one of the biggest INGOs in the world with an annual budget of over \$1 billion (Weiss, 2013). It receives funds from governments but the majority of its funding is from private donors (Dionne, 2014). World Vision is an example of an organization where private funders did call the organization to account for its policies and the threat to funding was enough to lead to change in policy. In March 2014 World Vision America, a branch of World Vision International, announced it would hire gay individuals in legally recognized same-sex marriages but reversed this decision three days later following negative reaction from conservative Christians. World Vision, and other religious organizations such as Christian Aid, Islamic Relief and Tear Fund implement programs providing services. There are also religious foundations that act as donors,

and some that operate both as donors and implementers, such as Caritas and Catholic Relief Services.

Research indicates corporations and religious donors will often follow similar practices to government donors in regards to accountability, reporting and evaluation (Chianca, 2008). This may be a tendency to follow what they see as industry standards that have already been set. It may be that if the influence of private donors continues to grow there may be a shift in the balance of power within the sector. However it seems unlikely that private corporations in particular will have particularly different goals than the national interest goals of Western donors.

Summary

This discussion shows the wide range of the positions of power various stakeholders in the humanitarian sector hold. Although each stakeholder does have access to certain means of power, this should not disguise the fact there are great power imbalances within the system. The humanitarian structure is set up in a way that allows donors, and in particular Western governments to hold considerable power over every other stakeholder in the system. In a simplified fashion, the structure currently dictates donors delegate responsibility for implementing programs to international organizations. These programs are supposed to be on behalf of target communities in the recipient countries, but the target communities have little ability to hold either donors or implementing organizations to account through answerability and even less with enforceability.

There have been developments that may change this dynamic in the future. Donors have become increasingly interested in participatory approaches and ensuring the views of beneficiaries are incorporated into programs. As such it is possible to envisage the donors

delegating power to the beneficiaries to hold INGOs to account for their actions, and systems being set up to ensure penalties for non-compliance with objectives. Additionally the increased involvement of private donors and non-Western government donors, will scramble existing power relationships. Opportunities to hold donors to account still remain limited, although some attempts via advocacy and the setting up of organizations dedicated to ensuring good behaviour by donors may improve these opportunities in future.

New Public Management's Influence on Accountability

Most of the discussion so far has concerned downward accountability. However the Asian Tsunami joint evaluation noted as a result of the large amounts of funding available for the tsunami response, INGOs had “a virtual obsession with ‘upward accountability’ to donors, the media and the public in donor’ countries” (Cosgrove, 2007, p. 11) and as the HAP surveys show it is still considered the dominant form of accountability in the humanitarian sector (Alexander, 2013). Alongside the debate about accountability to beneficiaries that stemmed from the Rwandan genocide, were questions of the effectiveness of aid and whether humanitarian relief produced positive outcomes cost-effectively. These questions were grounded in a neo-liberal new public management (NPM) approach to government that began to have an impact on humanitarian and development aid in the 1990s.

Although NPM began in the 1980s its impact in the humanitarian and development aid sectors became evident in the early 1990s as the concept of results based management (RBM) was introduced by most Western donor countries (Feinstein & Beck, 2006). Broadly defined, NPM is the use of private sector concepts of performance management, results oriented approaches and cost effectiveness in the public sector (Hood, 2001). NPM, it was argued, would modernize public sector service delivery by focusing on the public as a consumer within a free

market system that would improve public choice and thus quality. NPM evolved from the Reaganite and Thatcherite philosophies of smaller, decentralized government, which it was argued would ensure accountability and transparency to the tax-payer (Larner, 2000).

Performance management techniques that stressed quality and efficiency with a focus on demonstrating impact using measurable indicators are a particular area of NPM that affected both development and humanitarian relief. One of the effects of RBM was to shift the focus of aid planning and grant proposals from inputs to outputs and outcomes (Feinstein & Beck, 2006). The emphasis on measuring impact on clearly specified indicators was achieved at the expense of the ethical mandate of organizations: demonstrating outcomes were achieved became more important than serving beneficiaries and the participation of target communities suffered (Salgado, 2008). Focusing on measurable impacts may also reduce focus on hard to measure goals such as human rights protection (Hofmann, Roberts, Shoham, & Harvey, 2004). Therefore, the goal of NPM in humanitarian relief may be to improve quality but the effects can often be to reduce program delivery to limited indicators and reduce the involvement of, and thus accountability to, beneficiaries. One of the problems with the accountability that NPM seeks is that it intended to demonstrate efficiency, transparency and a lack of corruption to taxpayers, but in the case of humanitarian aid the taxpayers are not the recipients of the services provided thus diverting power affected communities have to hold both program implementers and donors to account.

NPM is primarily seen as establishing an upward form of accountability (Chouinard, 2013). However, a review of humanitarian aid standards and policies show the effect of NPM on initiatives that are part of the movement for accountability to affected communities. Many of the accountability standards reduce performance to intended beneficiaries to pre-determined

indicators (Everett & Friesen, 2010). For example, the Sphere Standards focus on technical aspects of aid delivery with a one size fits all prescription. Médecins Sans Frontières (MSF), one of the largest INGOs working in the sector, has been critical of this approach (Tong, 2004). MSF, and others, see the universal standardization of aid reduced to indicators and standards as harmful attempts to respond to the local context and learn from past programming mistakes (Walker & Maxwell, 2009).

HAP's initial standards also used language influenced by the NPM movement. Corporate terms and referrals to 'customers' and 'product and service value' (Everett & Friesen, 2010) highlight a business model approach similar to those espoused by NPM. Everett & Friesen (2010) note although HAP has reduced its reliance on the language of commerce in more recent publications, it still seems comfortable with allowing themselves to be associated with a corporate management model.

The HAP certification process also highlights an interesting tension in the accountability debate. To become certified by HAP a member has to go through a certification process demonstrating to HAP its commitment to accountability to the communities they work with and quality management (HAP, 2014). This process requires a certification process that requires a program audit at the organization's headquarters and at least one field site. This is just one of many certification programs that exist. Others include International Organization for Standardization (ISO 9000) certification and the Société Générale de Surveillance (SGS) benchmarking system. In an increasingly crowded field of emergency response the certification process is a method for signaling quality and an organization's commitment to being accountable to its intended beneficiaries. However to gain certification, INGOs have to make themselves accountable to another organization and not their intended beneficiaries. In effect they are

proving they are accountable to intended beneficiaries by being accountable to an international organization.

Evaluation Within the Humanitarian Sector

Evaluation is increasingly seen as a key tool in demonstrating accountability in the humanitarian sector. Despite the multitude of policies, methodologies and driving forces in the evaluation process in the international aid sector, there is one overarching framework from which most donor and NGO evaluation policies are developed—the 1991 Development Assistance Committee Evaluation Guidelines developed for the OECD (Chianca, 2008). The Committee consisted of powerful players in the development community including the heads of most of the bilateral and multilateral donor agencies (Chianca, 2008). The development guidelines are important as they formed the basis for guidelines on emergency response. The guidelines had six basic principles:

- “1. All aid agencies should have an evaluation policy.
- 2. Evaluations should be impartial and independent.
- 3. Evaluation results should be widely disseminated.
- 4. Evaluation should be used—feedback to decision-makers is essential.
- 5. Donor and recipient agencies should be partners/cooperate with the evaluation—strengthen recipient agencies and reduce administrative burden.
- 6. Evaluation should be part of the aid planning from the start—clear objectives are essential for an objective evaluation.” (Chianca, 2008, p. 42)

OECD/DAC has defined development evaluation as the “systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives,

development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. Evaluation also refers to the process of determining the worth or significance of an activity, policy or program.” (OECD DAC, 2010, p.

4) Although developed for international development evaluation rather than EHA this definition is none-the-less relevant because it is often used by donors and INGOs in their evaluation policies.

The language of the six principles emphasizes upward accountability. Guideline 2 emphasizes the importance of independence of the evaluation, something also emphasized in the 2010 guidelines (OECD DAC, 2010). Whilst independent evaluation does not prevent the possibility of downward accountability, it does reduce options for participatory approaches and thus suggests upward accountability was the main consideration of the guidelines. Guideline 5 considers donor and partner cooperation but does not refer to intended beneficiaries and guideline 4 refers to the importance of sharing results with decision makers. Both of these guidelines are heavily weighted towards upper level management and upward accountability. The only guideline that includes downward accountability is guideline 3, which calls for the wide-dissemination of results. However the broadness of this guideline does not emphasize dissemination to intended beneficiaries and can easily be interpreted as just lesson learning among other agencies.

The five criteria highlighted in the OECD’s definition of evaluation (relevance, efficiency, effectiveness, sustainability, and impact) were a key part of the original 1991 document. In 1999, the DAC released a document addressing complex emergency programs. The guidelines and criteria were similar to the 1991 document but were altered to reflect the unique

circumstances of emergency assistance. The criteria were expanded to relevance/appropriateness, connectedness, coherence, coverage, efficiency, effectiveness, and impact. Today, these criteria are present in some form in most donor and NGO evaluation policies (Chianca, 2008).

Another widespread definition of evaluation of humanitarian action (EHA) is one used by the ALNAP. ALNAP bases its evaluation guidelines on the 1999 DAC/OECD humanitarian relief guidelines and criteria. ALNAP (2006) defines evaluation as follows.

“Evaluation of humanitarian action (EHA) is a systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability.

EHA:

- is commissioned by or in cooperation with the organisation(s) whose performance is being evaluated;
- 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;
- assesses policy and/or practice against recognised criteria (eg, the DAC criteria);
- articulates findings, draws conclusions and makes recommendations.”(p.14)

The DAC criteria give more weight to broader definition of accountability than the six OECD guidelines. ALNAP (2006) has produced a widely-used guide to EHA that defines each criterion and highlights types of evaluations relevant to particular criteria. The types listed include single agency, multi-agency, sector wide, multi-sector, institutionally focused, policy focused and partnership focused evaluations (ALNAP, 2006, pp. 21–22). The language used in the guide suggests both upward and downward accountability issues are important. For example,

the explanatory notes on the criterion of efficiency focus on the technical aspects of value for money and suggests an economist be included in the evaluation team. The emphasis is on what outputs were achieved with the inputs, with no reference to long-term outcomes. In this section the guide uses language associated with NPM's approach to efficiency. The section on relevance/appropriateness focuses on local needs and increased ownership by the target population, suggesting more of an emphasis on downward accountability. Additionally, by raising questions of coherence and coverage, the guide also addresses the broader political and economic context of the intervention. As such the guide provides a much broader approach to questions of accountability than the six guidelines of the DAC/OECD document.

ALNAP's guide notes all seven criteria may be useful in an evaluation, but most INGO evaluation policies do not distinguish among the seven criteria (in fact many agencies have the same policy for their development work as their humanitarian work, and only list the original five criteria).

In many ways ALNAP's evaluation guidelines and the DAC/OECD criteria have the same impact of universalization HAP's accountability certification or the Sphere Standards have (Walker & Maxwell, 2009). These documents have become the dominant scripts in EHA. Whether this dominance in INGOs evaluation policies creates flexibility that allows evaluators to apply the criteria effectively to a local context or instead creates rigidity that harms the overall utility of an evaluation is an important empirical question.

EHA and Accountability

One of the key distinctions between evaluation and research is evaluation places a value judgment on something (Mathison, 2008; Weiss, 1972). Evaluating something does not

necessarily lead to accountability, but the process of accountability does involve making a value judgment and hence requires evaluation (Mathison, 2009). Ebrahim (2003) identifies five categories of accountability mechanisms; evaluation, reports and disclosure statements, participation, self-regulation, and social audits. As such there is no question evaluation can play a role in accountability. The question is more what type of accountability can and does evaluation contribute to?

There is considerable debate within the sector as to how accountability can be addressed through evaluation (Hallam, 2011). Feinstein & Beck (2006) argue in the early stages of the accountability revolution there was a belief evaluations could either focus on lesson learning or accountability, and there was an inherent trade-off between the two. This has been challenged more recently by the concept of accountability for learning which argues the concepts can play a complementary function in evaluation.

The debate about accountability versus lesson learning extends to consideration of what evaluation approaches should be used. The first draft of ALNAP's guide for EHA states evaluations play either a lesson learning role or accountability role, and argues an evaluation mainly focused on accountability is not likely to use participatory evaluation approaches (ALNAP, 2003). However, by the time the final version was published in 2006, after field testing and feedback, this distinction has been removed (ALNAP, 2006), which supports Feinstein and Beck's view the distinction was less clear cut than had previously been the case. The most recent ALNAP guide though states again accountability focused evaluations “are more likely to adopt a more adversarial investigative style...participatory and facilitated evaluations are a more appropriate approach when learning is the primary goal” (Cosgrove & Buchanan-Smith, 2014, p. 45)

The 2003 draft suggests the definition of accountability focused strongly on upward accountability. Indeed the draft guide quotes the OCED guidelines for complex emergencies: “A critical question to be considered at the outset is whether the evaluation is going to emphasize lesson learning or accountability, or a mix of the two. If lesson-learning is emphasized then it opens up the possibility for the extensive use of participatory methods. If accountability is emphasized then it implies structuring the evaluation so its findings are independent and respected.” (1999, p. 17). That the document that has had the most powerful effect on humanitarian aid evaluation work in the past 15 years states this, gives some indication of the power imbalances within the humanitarian sector. This suggests at the time of formulation, the committee considered accountability to be upward and focused on technical and contractual compliance. The working group that established the initial development guidelines consisted of representatives from 16 donor organizations and 6 UN agencies, and only 5 NGO representatives along with 1 representative from ICRC and 1 from IFRC (Chianca, 2008). There was therefore a strong donor presence in the formulation of the guidelines. The OECD itself is a Western organization with members that at the time of the guidelines being written were heavily involved in NPM reforms (Haque, 2007).

Summary

This review of literature suggests tensions exist between the different conceptions of accountability within humanitarian relief. On the one hand downward accountability receives much of the attention in literature and policy debates with continued arguments as to how much progress has been made. On the other hand, what evidence exists suggest upward accountability driven by a need to account to powerful actors who control funds is still the dominant form of accountability. Debate over how evaluation can be used as an accountability mechanism is also

ongoing. Building on these continued debates, this study investigated how and whether evaluation is being used as a downward accountability mechanism, and who benefits from current evaluation practices in humanitarian aid.

Evaluation Approaches

There are numerous evaluation theories, models and approaches. Evaluators have differing views as to how to approach evaluations and the organizations are likely to seek out evaluators whose approach matches their ideas for the purpose of evaluation. Evaluators may see their role as being important for social justice, to focus on organization change, or place a value or judgment on a program. Broadly speaking, evaluators focus on use, values or methods, and would have been influenced by the works of theorists working in the evaluation field such as those identified by Alkin in his evaluation theory tree (Alkin, 2004; Christie & Alkin, 2008). Some evaluation approaches are more common in EHA. Some of these approaches are most relevant to downward accountability, whilst the use of other approaches would suggest more of a focus on upward accountability. There are also approaches where either upward or downward accountability may be favoured dependent on how the evaluation is actually implemented.

Evaluation Approaches Favouring Upward Accountability

Evaluations favouring upward accountability are more likely to focus on technical and theoretical approaches. The evaluations would be expected to highlight the important elements of NPM, namely a focus on performance management and achieving of measureable indicators. The approaches are also more likely to favour comparative measurement of impact through the use of experimental or quasi-experimental designs and a strong counter-factual. The NPM movement has impacted accountability policies domestically in many Western countries.

Governments in both Canada and the United States have developed policies that favour so-called ‘more scientific’ approaches (Chouinard, 2013).

Experimental and quasi-experimental designs. Experimental and quasi-experimental designs are grounded in positivist models of inquiry that require a control group to allow a researcher to analyze difference between groups. In humanitarian aid this can prove both difficult and morally suspect. Humanitarian aid is supposed to be given on the basis of need and so denying a group support to allow for the measurement of impact on another group raises considerable moral questions. Quasi-experimental design involves the non-random selection of the control group, which when appropriately chosen may allow the researcher to create a control group similar in characteristics to the population impacted by the project.

Theory based approaches. The purpose of a theory based approach to evaluation is answer what is causing an approach to work, as well as the actual performance or merit of the intervention (Chen, 2005; DfID, 2013). This approach relies on identifying either the theory of change or the logical model behind the intervention and then testing if the theory the intervention was based on holds true. Theory based approaches are often associated with experimental or quasi experimental design (Hacsi, 2000). Theory based approaches have also been linked to a concentration on achieving indicators rather than the long-term viability of the program when used for accountability purposes (Rogers, Petrosino, Huebner, Tracey, & Hacsi, 2000), a criticism which is similar to criticisms of NPM approaches.

However, a theory based approach to evaluation does not automatically assume the accountability direction is upward. Participatory approaches can be used to identify and test the assumptions of the theory of intervention. This will rely on beneficiaries being involved in the initial design of the theory and the evaluation questions that are needed to test it.

Impact evaluation. Impact evaluation has been increasingly used in the international development context. The goal of an impact evaluation is to measure the long-term effects of a project or intervention. Governments are increasingly interested in measuring the impact of humanitarian relief projects. For example in a recent policy document, the UK's Department for International Development (DfID), promises to “invest more in measuring the UK Government’s impact and the impact of our partners” (2011, p. 16). The evaluation guidelines of major donors argue impact evaluation requires a counter-factual and the use of statistical data is important to identifying causation (DfID, 2011; USAID, 2013). The United States Agency for International Development (USAID) specifies experimental or quasi-experimental designs are the only appropriate designs in the evaluation. This approach is not shared by all donors; DfID for example highlights the importance of qualitative data to give a contextual emphasis to findings. However DfID still focuses more on experimental design in its guidelines than qualitative methods. Therefore, similarly to theory-based evaluations, impact evaluations do not have to be focused on upward accountability, but the current guidelines issued by leading international donors heavily leans towards this with the emphasis on experimental design.

Technical evaluations. Evaluations where the goal is focused on identifying the technical quality of a program are also often more aimed at upward accountability than downward accountability. As an example an evaluation looking at the quality of prosthetic and orthotic (P&O) services after an earthquake may involve a technical expert with limited local knowledge. The evaluation may focus on cost-effectiveness and the quality of the P&O services compared to international standards. While this does not preclude downward accountability, local need and the effects of the local environment on the P&O devices may be forgotten. In this case the evaluation focuses on demonstrating to donors, headquarter quality control specialists

and quality monitoring bodies that the program has met certain quality and cost-effectiveness goals. This hypothetical example highlights a common criticism of technical evaluations that focus too heavily on technical data and not enough on local context (Riddell, 2007).

Evaluation Approaches Favouring Downward Accountability

Evaluation approaches favouring downward accountability focus on the intended beneficiaries and target communities of the program. As such, approaches involving participation and collaboration are likely to be most appropriate for this form of accountability. These approaches are a subset of stakeholder evaluation approaches, but ones that focus most particularly on one stakeholder—the intended beneficiaries. Approaches particularly relevant here are participatory evaluation, empowerment evaluation, equity-focused evaluation and collaborative evaluation. There are also broad approaches that can be an umbrella approach, such as utilization-focused evaluation (UFE). For example, it is possible for UFE to emphasize ‘participation’ as central to the evaluation.

Participatory evaluation (PE). Participatory approaches developed in 1970s and 1980s in the international development sector and were the starting point for collaborative approaches in evaluation (Cousins, Whitmore, & Shulha, 2013). Development work in Latin America influenced the development of PE (Cousins et al., 2013; Kushner & Rotondo, 2012), and was based on participatory tools such as rapid rural appraisals and participatory action research (Cousins et al., 2013). These approaches developed as a reaction to more established positivist evaluation approaches that were considered exploitative (Cousins et al., 2013). Participatory evaluation advocates for the joint control of the process by the stakeholders and the evaluator (Cousins & Whitmore, 1998; Fetterman, Rodriguez-Campos, Wandersman, & O’Sullivan, 2013). Cousins & Whitmore (1998) identified two forms of participatory evaluation; practical

participatory evaluation (P-PE) and transformational participatory evaluation (T-PE). P-PE is grounded in a utilization-focused approach based in the belief stakeholder involvement in evaluation improves the chances the findings will be used for program improvement. This occurs because stakeholder involvement improves the relevance and ownership of the evaluation (Cousins & Chouinard, 2012; Cousins & Whitmore, 1998). T-PE is based on the concept of knowledge ownership and is closely linked to the 1970s movements in Latin America. It seeks empowerment and social justice through ensuring program participants are active partners and owners of the knowledge produced about that program (Cousins & Whitmore, 1998).

Empowerment evaluation. Empowerment evaluation was conceptualized by David Fetterman in the early 1990s. Empowerment evaluation focuses on self-determination and is grounded in the ideas of community psychology and action anthropology (Fetterman, 1994, 2005). The approach was met with resistance from some of the evaluation community but has still formed an integral part of participatory evaluation approaches (Cousins et al., 2013). As an approach, it puts evaluation in the hands of the local community and program staff with the evaluator playing a guiding role.

There are ten principles of empowerment evaluation: improvement; community ownership; inclusion; democratic participation; social justice; community knowledge; evidence-based strategies; capacity building; organizational learning; and, accountability (Fetterman, 2010). “The accountability principle guides community members to hold one another accountable. It also places the evaluation within the context of external requirements. The community is accountable for reaching specific standards or delivering specific results, products and outcomes” (Fetterman, 2010, pp. 280–281). This idea pays attention to both upward and downward accountability but does place the responsibility for accountability on the community

as a whole, not just the organization delivering the program. Fetterman considers program staff to be part of the community so there is still a role for the organization to play.

Collaborative evaluation. Collaborative evaluation shares many common features with participatory and empowerment evaluation approaches. In collaborative evaluation, evaluators are more in control of the evaluation than participatory and empowerment approaches, but the importance of creating a strong relationship between evaluators and stakeholders is stressed strongly (Fetterman et al., 2013; Rodríguez-Campos, 2012). Collaborative evaluation is therefore defined as ‘an evaluation in which there is a substantial degree of collaboration between evaluators and stakeholders in the evaluation process, to the extent to which they are willing and capable of being involved.’ (Rodriguez-Campos & Rincones-Gomez, 2012, p.4). This definition is important as it suggests stakeholders may not always wish to be involved. This seems potentially relevant to EHA in contexts where the ongoing fight for survival and recovery could easily make participating in an evaluation seem irrelevant to a community member. Understanding the context of the evaluation is important to distinguish whether local sentiments are that an evaluation is not relevant because of a general belief INGOs do not listen and so the evaluation serves no purpose, or other issues that make involvement difficult.

Utilization-focused evaluation (UFE). UFE developed in the 1970s and 80s as evaluators began to notice there was limited evidence evaluations were being used (Patton, 1997). Michael Quinn Patton is widely accepted as the leading proponent of UFE (Patton, 1978; Patton, 2008; Patton, 2012). As is evident in the name, the underlying idea is evaluations should be used. As such, an evaluator is responsible for approaching all aspects of the evaluation with the consideration of use in mind. This requires identifying potential users and working to ensure they are key participants in the evaluation (Patton, 1997).

UFE is broader approach than participatory or empowerment evaluation, and is not necessarily an approach that focuses on the affected community, however, it is an approach both participatory and empowerment approaches would fit within. If the evaluation identifies the anticipated users as the intended community, then the evaluation can be designed to ensure the use of the evaluation by this group for the purpose of accountability. Patton argues UFE is “inherently participatory and collaborative in actively involving primary intended users in all aspects of the evaluation” (1997, p. 100), thus linking the approach most closely to P-PE. In this approach, intended use is closely linked to organizational learning and thus an approach that makes sense for those who argue lesson learning and accountability are compatible in the same evaluation.

Deliberative democratic and democratic evaluation. The ideas of both Ernest House and Barry MacDonald focus on the use of democratic principles to provide evaluations to strengthen social justice. MacDonald identifies three different categories of evaluation, bureaucratic, autocratic and democratic (MacDonald & Kushner, 2005). MacDonald’s initial concern was the power to make decisions about education policy was passing from professional educators to civil servants and politicians (MacDonald & Kushner, 2005). Democratic evaluation offered a means to protect against this by the evaluator ensuring multiple viewpoints are included in the evaluation. In this respect the evaluator’s role is that of a facilitator. Although his initial work focused on education, MacDonald’s work is relevant to other contexts including humanitarian relief, and it is particularly relevant in discussions about accountability as he frames his ideas around the rights of access to information.

House’s work on democratic deliberative evaluation is likewise an approach that can support downward accountability. For House the role of the evaluator is to try to understand and

include the views of a multitude of stakeholders, and in particular those that are marginalized and powerless (House, 1980).

Real-time evaluation (RTE). Real-time evaluation has become increasingly used in the humanitarian sector. The purpose of RTE is to give immediate, on the ground feedback to an ongoing project/program. It is an approach that recognizes the urgency of many humanitarian responses, and that feedback from an end of project evaluation may come too late to help the intended beneficiaries of a particular invention. The primary intended users of RTEs are field staff with the goal of improving field operations in an ongoing emergency (Cosgrove, Beck, & Ramalingam, 2009). In RTEs, evaluators present reports before they leave the field site where evidence is collected, and thus is a formative evaluation that aims at instant impact on program operations.

The evolution of the debate about whether accountability and lesson learning are compatible can be seen in the literature on RTE. Herson and Mitchell (2005) argue RTE is not strongly aimed at accountability and its main focus is lesson learning, although they do acknowledge RTEs may promote downward accountability. Polastro (2011) argues because RTEs are participatory in nature, they can play a crucial role in improving accountability to intended beneficiaries. He does however acknowledge while this should be the case, in many evaluations, RTEs are still headquarter driven, with a focus on accountability to donors. RTE is another example of an evaluation approach that could fit either upward or downward accountability. The intended goals of the evaluation, the data collection methods used, and the anticipated use of the findings will impact what accountability is being addressed.

The Quality and Intent of Evaluation

The type of evaluation approach may not matter if the quality of the evaluation is poor and the intent is just to tick the ‘evaluation completed’ box. Considerable criticism has been made about the quality of evaluations within the humanitarian relief sector (Feinstein & Beck, 2006; Gibson, Andersson, Ostrom, & Shivakumar, 2006; Pérouse de Montclos, 2012; Riddell, 2007). These criticisms include that many evaluations have weak TORs, do not adequately consult stakeholders, pay limited attention to future use, and have poor methodological standards (Feinstein & Beck, 2006). Another further weakness is the misuse of terminology, in particular the term ‘participatory’. This builds the impression evaluations are being done because they are required either by donors or INGO policy, but limited effort is being put into them. In these cases, the approach and methodology is irrelevant. A poor quality evaluation, with limited enthusiasm from the INGO’s field staff is unlikely to provide downward accountability to beneficiaries.

The Logic of Evaluation

With many extant evaluation approaches, one might ask, ‘what is the purpose of the evaluation?’ At the heart of this question lies the logic of evaluation. Scriven’s (2005) definition of evaluation is one of the most important contributions to the idea of the logic of evaluation he defines evaluation as the determination of the merit, worth or significance of a social intervention. Who makes these determinations, against what criteria, and to what audience influences the evaluation approach selected. House (1980) argues evaluation is about persuasion aimed at particular audiences: “persuasion claims validity only for particular audiences and the intensity with which particular audiences accept the evaluative findings is a measure of their effectiveness” (p. 73). Acknowledging an evaluation cannot reach all audiences, the approaches

outlined above, take different approaches to whom the evaluation will reach and how they will persuade the audiences of their findings.

Deciding what criteria and standards will be used for an evaluation will often be determined by the approach used and heavily influenced by the target audience, and in the context of EHA, the type of accountability desired. Although the OECD/DAC guidelines have led to a generally unified overall approach to criteria (Chianca, 2008), within this framework there is considerable flexibility in the weight given to a particular criterion, the development of sub-criteria, and what standards are applied.

Evaluative claims are the outcomes of doing an evaluation, and an evaluator uses warrants to justify those claims. Warrants allow evaluators to make claims based on the evidence they have found by relying on some kind of authority (Fournier, 2005). Warrants will vary depending upon the audience. For example, an evaluation emphasizing the criterion of effectiveness in a disaster reduction project may use various engineering tests to demonstrate effectiveness and engineering industry standards might be used to consider the evidence of effectiveness. As such, the evidence to support the claim of effectiveness is the data of an engineering test compared against industry standards. This evidence may mean nothing to the rural community facing a cyclone on the coast of Bangladesh. For them, effectiveness may mean how many lives were saved when the cyclone hit, and thus the relevant evidence for this group is the stories of community members who used the shelter or perhaps a comparison of the number of casualties in this cyclone compared to previous cyclones. How evaluations treat questions of persuasion, who is involved in developing criteria and standards, and how claims are presented based on the evidence found, all impact the accountability focus of the evaluation.

Even with approaches that focus on participation and empowerment, the evaluator, or the commissioner of the evaluation is still making choices that exclude certain stakeholders or groups (House, 1980). It is impossible to represent every interest or group in a community project evaluation, and less so in a region or country-wide evaluation. Better organized groups will tend to be better represented, both in needs assessments prior to a project and in an evaluation. One section of the community may also be better represented as issues affecting them are more pertinent in the eyes of donors, NGOs, society or the local community. House (1980) argues participatory evaluation is the most effective form of evaluation in including a wide group of diverse stakeholders, but even this approach has not completely escaped the problem of favouring one group over another.

Chapter 3: Research Methodology

Research Questions

The research questions guiding this study focus on understanding the current practices of EHA within the historical context of the accountability debate in the humanitarian sector.

The main research question is:

How are INGOs operating in humanitarian relief contexts using evaluation as a mechanism to provide and/or assess accountability to affected populations?

Sub questions investigated are:

What particular evaluation approaches are being used with which forms of accountability in humanitarian relief work?

Who benefits from current evaluation practices?

The wording of the questions originally was constructed to investigate accountability to intended beneficiaries. This was broadened to affected populations to reflect that programs and projects by INGOs impact entire communities, and not just the direct recipients of the program.

The Listening Project identified beneficiary selection criteria as being one of the main areas where communities feel there is a lack of information from INGOs (Anderson, Brown, & Jean, 2012). Providing information to a broad section of the community is important to reduce knowledge gaps. As such my research looked at the inclusion of the community and not just beneficiaries in the evaluation. This also reflects the suggestions of Scriven (2007) to avoid the word beneficiary as it pre-supposes those affected by the intervention actually benefitted from it. This is not automatically the case.

Researcher Background

My background, experiences and perspectives play a role in identifying what I consider to be an interesting research problem and reasonable ways to engage in social science inquiry. The importance of describing researcher experience is grounded within the social constructionist approaches that recognize researchers' interpretations reflect knowledge, values and interests researchers bring to their studies. As such, presenting what Van Maanen (1988) calls a confessional tale allows me to situate myself within the research.

I have worked in international aid for much of the last 10 years, including placements in Tanzania, Thailand, Indonesia, Sri Lanka, Afghanistan and Bangladesh. This work has been split between international development and humanitarian relief. I have worked for organizations most closely aligned with what Walker and Maxwell (2009) describe as the "solidarist" group of humanitarian relief NGOs. The solidarist group has an expanded definition of humanitarian relief beyond the narrow traditional, minimalist version espoused by organizations such as MSF. This includes humanitarian relief work focused on more than immediate life-saving needs and looks at human rights and social transformation as well. I have worked in long-term refugee situations with Karen refugees on the Thai-Burma border and a controversial political environment in Afghanistan. A lot of my work has been in the fuzzy area between humanitarian relief and development. I have also worked in contexts that clearly fit into the development definition but were funded with remaining funds from an emergency appeal. A road-building project on the Indonesian island of Nias, which was funded through money raised by CARITAS in response to a large earthquake in 2005, is an example of this.

I have participated in evaluations in the field and understand how evaluation is embedded in project cycle management practices. In general, my experience has been evaluation is often an

afterthought in both humanitarian relief and international development projects, and seen as a burden mainly aimed at pleasing donors. The monitoring and evaluation systems I have seen implemented seem designed to count project indicators for donor reports and to provide counts of the number of people reached, which becomes a key tool for head offices to advertise the work they are doing to donors, the media and the public in the organization's home country. The idea of downward accountability has been alien in these evaluations. Downward accountability fares better in project design and ongoing communication during a project. Mechanisms to listen to beneficiaries during the project design and ensure their participation in the project are often in place, although the effectiveness of this varies from project to project. My experience has been that generally at all stages in the project cycle the inclusion of, and accountability to, beneficiaries is dependent upon the attitude of individual field staff. NGO head offices put pressure on country offices for upward accountability and implement policies that pay lip service to downward accountability, but how these policies are implemented is generally left to the field offices. My work is limited to three different INGOs, none of which are the very large INGOs such as CARE International or Save the Children.

Critical Hermeneutics

The theoretical and methodological framework for this study was critical hermeneutics. Evaluation reports are used as an organizational tool to communicate information and knowledge about a INGOs performance on a project or program. They are texts produced within the organizational culture of the INGO and sectorial culture of the humanitarian community. Within the sector, the ideas of accountability and the use of evaluation are developed through the practices of individuals and organizations. The evaluation reports are developed, understood and utilized by practitioners and stakeholders through a system of interpretation socially constructed

by common industry practice. Not all stakeholders have had equal input into the development of these practices and therefore an understanding how power intersects with the interpretation of text is important to recognizing how evaluation reports contribute to, maintain, or seek to alter the structure of the humanitarian sector. Critical hermeneutics offers a framework for understanding the relationship among evaluation reports, accountability and power. A critical hermeneutic approach will go beyond the evaluation reports themselves to consider the cultural and historical context in which these documents were produced and supports the questioning of the social and political conditions that led to the construction of the current structure in the humanitarian sector.

The roots of hermeneutics are in the work of German philosophers and is a framework for interpreting text (Crotty, 1998; Myers, 1995). Historically, the context in which the text was written was important in interpreting the authors' meaning. Today, hermeneutics is used in various social science disciplines to understand cultures, organizations and society (Crotty, 1998; Myers, 1995; von Zweck, Paterson, & Pentland, 2008) and de-emphasizes the authors' meaning. A hermeneutic approach is grounded in the idea that an interpretation of documents is just that, an interpretation. A different researcher could well have a different interpretation and the researcher's experiences and perspectives are a crucial part of the interpretation (Patton, 2002). This was important for this research as it allowed the use of my own knowledge and experience of humanitarian aid and the practice of evaluation as part of the interpretative process. This experience helped facilitate interpreting data from evaluation reports and interviews to create a new interpretation of practice of evaluation.

Kinsella (2006) offers five characteristics of the hermeneutic methodology: "(a) seeks understanding rather than explanation; (b) acknowledges the situated location of interpretation;

(c) recognizes the role of language and historicity in interpretation; (d) views inquiry as conversation; and (e) is comfortable with ambiguity” (p. 3). These characteristics demonstrate why hermeneutics was a useful approach to this research. My research aimed to understand the context and history of the accountability revolution within the humanitarian relief context and how this has impacted the use of evaluation. The hermeneutic approach recognizes that interpretation is based on the situated viewpoint of the researcher, thus allowing the researcher’s knowledge and experience to fuse with new perspectives gained from the research (von Zweck et al., 2008). Therefore a hermeneutic approach allowed me to recognize and use my own experiences for understanding evaluation practice. The historical perspective was also necessary for understanding issues of power and dominance that led to the development of current evaluation practices.

A criticism of hermeneutics is it takes texts at face value (Myers, 1995). In contrast, critical hermeneutics expands the interpretative frame to acknowledge issues of power, dominance and suppression within the contexts in which a text was produced. This approach recognizes documents are ideological texts and grounds interpretation within the power structures in which the texts were produced (Roberge, 2011). Critical hermeneutics developed from the dispute between Hans Gadamer and Jurgen Habermas during the 1960s and 1970s over the idea of universal hermeneutics and critical theory (Roberge, 2011). The consequence of this debate was the development of an approach that fused the interpretative elements of Gadamer’s hermeneutics with the critical theories of Habermas (Myers, 1995). Gadamer recognized human experience must be understood within a historical context (Roberge, 2011) and developed a theory of dialectical hermeneutics incorporating the traditional ideas of the hermeneutic circle

with an understanding that historical context is understood from the position and experiences of the researcher (Myers, 1995).

Paul Ricoeur has been credited with helping to develop an understanding of dialectical hermeneutics that highlights the critical potential of Gadamer's work (Myers, 1995; Roberge, 2011). Myers (1995) argues a critical hermeneutical approach allows the researcher to critically analyze an individual's or organization's understanding of an event within the terms of the structures which they operate in. This ensures imbalances of power are recognized in the researcher's analysis. In this research, the actions of evaluators, program staff and other stakeholders were analyzed within the structural organization of the humanitarian sector. This approach allows the researcher to focus on the power imbalances within the sector and interpret reports and interviews accordingly.

Roberge (2011) argues a full understanding of critical hermeneutics requires a third element that completes the Habermas and Gadamer debate over the meaning of hermeneutics, and the importance of experience and historical context. This is a theory that interprets action as "opposing performances driven by ideological-moral views" (p.1). His theory of action sees text as a performance where text is used by actors to persuade others of the dominance, righteousness, or authenticity of their position. Roberge argues Ricoeur's conception of ideology as text has always relied on interpretation, and ideology is used as a means of control, retention of power, and as opposition to power. Both those who hold power and those who oppose that power use ideology in the manner Ricoeur suggests. Roberge's theory of action is relevant to the investigation of evaluation as an accountability mechanism if one frames the push for downward accountability as a social movement and the focus on upward accountability to donors as a counter-movement. Social movements and counter movements create the dialectic of struggle

against each other where social reality is interpreted according to one's ideology. This struggle dovetails with the theory of experience both Roberge and Kinsella (2006) identify as central to interpretation. For the purpose of this study, the experiences of organizations and the historical understanding of accountability and evaluation are critical to how both are practiced today. INGOs are participating in a social movement focused on increased accountability to beneficiaries and the counter movement that prioritizes accountability to donors.

Summary

Critical hermeneutics therefore offers a framework for investigating how the understanding of accountability and the use of evaluation have contributed to building the structure of the humanitarian industry as it is today. The approach allowed me to ask how evaluation reports and other text, such as evaluation policies, contribute to maintaining or changing the status quo in the industry. It can show how individuals and organizations understand key language and how this governs the interaction between key stakeholders.

This research study used critical hermeneutics to analyze the meaning of evaluation in the humanitarian sector. The foundational question in hermeneutics is “what are the conditions under which a human act took place or a product was produced that make it possible to interpret its meanings?” (Patton, 2002, p. 113). The research questions take this approach with the added use of dialectics to ask how we can understand EHA within historical and current power relations when applied to the accountability movement. Dialectics was used to examine the dominant ideologies of accountability in the sector and asking how these ideologies affect understanding and experience of evaluation, and what benefits and exclusions current practice creates.

Data Collection Methods

Two primary sources of data were used to answer the research questions: 1) published evaluation reports and 2) semi-structured interviews with evaluators and NGO staff. Secondary data included NGO evaluation policies and accountability statements, and various accountability initiative documents. How evaluation reports and interviewees were selected and how data collected is described below.

Evaluation Reports

Sample selection. Evaluation reports were selected from reports published on the ALNAP website in a multi-stage process. The reports were found by following the link on ALNAP's home page called “evaluative resources only”. This page allows search criteria to be applied to their list of resources. The filters used are resource type, date and language. For these groups the selection criteria are:

- Evaluation Reports (from resource type filter)
- English (from language filter)
- January 2012-March 2014 (from date filter)

The date filter was applied to ensure the review was reflective of current evaluation practices.

The language abilities of the researcher limited the sample to English reports only.

A total of 238 documents were housed on the ALNAP website once filters are applied to the search engine when the initial selection. The second stage of the process was conducted by review of the title of the report and where necessary the summary page of the evaluation. For example, under the selection criteria it was easy to exclude a report titled “Meta-Evaluation of Quality and Coverage of USAID Evaluations” based on the title, but a report with the name

“Emergency Operation of Ethiopia drought operation” required closer review to recognize this was published by the International Federation of the Red Cross and thus met one of the exclusion criteria.

The inclusion criteria at this stage of the sample selection were:

- Single NGO single project/program reports
- Multiple NGO single project/program reports
- Joint NGO and UN projects/programs where a NGO was the lead organization

The exclusion criteria included:

- UN agency reports
- ICRC, IFRC, and other Red Cross reports
- Donor led reports
- Sector-wide reports
- Evaluations that only published the executive summary or management response
- Emergency funding umbrella organization’s evaluations
- Advocacy and research documents

The document search still contained a small number of documents that were not evaluation reports, presumably misclassified on ALNAP’s system. These included advocacy reports by NGOs and research reports published by universities. In particular, documents posted in 2014 included a number of advocacy reports about the drastic situation facing children in Syria. These were removed from the sample. There were also a number of double postings. Oxfam for example will often post the initial report and then add another posting which includes other documents such as the management response as well, while leaving the original posting on

the site. The remaining documents could be split into broad categories. These were single NGO, multiple NGOs, UN agency, Red Cross, donor, sector-wide reports, and evaluations of projects from a specific fund-raising appeal.

The UN agencies', Red Cross and donor evaluations were excluded from the sample. The research interest is on INGOs. Although the practices of the UN and Red Cross community are important topics for research, in this case in order to maximize the number of NGO evaluations that could be analyzed a decision was taken to exclude UN and Red Cross evaluations from the sample. There are differences in power dynamics involving the UN and the Red Cross. Funding practices and status in country impact these dynamics. Additionally often budgets available for evaluations mean practices may be distinct. A further research project to look at the differences and similarities is necessary but is beyond the scope of this research.

A number of the evaluations on the ALNAP site were commissioned by NGO umbrella groups. These groups are often set up to coordinate fund-raising. An example is the Disaster Emergencies Committee (DEC) which was set up in 1963 to take advantage of an offer of free television advertising that was conditional on British NGOs responding to disasters making a general appeal rather than competing with one another for funds (Walker & Maxwell, 2009). DEC had a number of evaluations posted on the ALNAP website which fit into the other search filters. These were excluded from selection for two reasons. These evaluations cover a large number of NGOs and often a wide range of diverse projects, and as such are not in-depth. Additionally, many NGO projects had published evaluations on their individual agency's project that was part of the DAC (or other funding appeal body's) evaluation. Based on these factors a decision was taken to exclude these evaluations from the sample.

After applying the selection criteria, 46 evaluation reports were identified on the ALNAP website. The final sample was 10 evaluations reports, a number that allowed for in-depth analysis of the evaluation reports as text and also permits the inclusion of a variety of different contexts and INGOs. The selection criteria for the final sample included:

- Five reports from the five biggest INGOs; World Vision, MSF, Oxfam, Save the Children, and Care International.
- Five reports from the other NGOs not included in the list of the largest five.
- Reports that included a TOR (or for which a TOR can be found).
- Reports that included the name and preferably contact details of the evaluator, evaluation team, and program/project field staff focal point.

The initial plan of selecting one evaluation report from each of the five biggest INGOs was not feasible because MSF and World Vision do not published their evaluation reports (although World Vision will make their reports available on request). The sample therefore included 2 Oxfam evaluations and 2 CARE International evaluations. The INGOs represented in the final sample were Oxfam, CARE International, Save the Children, Support to Life, the Norwegian Refugee Council (NRC), the Danish Refugee Council (DRC), Christian Aid and Action Contre la Faim (ACF). The evaluations covered projects and programs in Haiti, Gaza, Mali, Niger, Chad, Uganda, Somalia, Kenya, Turkey, Pakistan, South Sudan and the Philippines. The interventions were in crisis that were both natural, such as earthquakes, floods and typhoons, and man-made, including war, refugee crises and occupation, and also crises caused by a combination of both, particularly famine stimulated by conflict. The evaluations also covered crises which were both immediate emergencies, such as the response to the Typhoon Haiyan in

the Philippines, and longer term crises, such as the occupation in Gaza. These longer term crises often fall into the greyer area between traditional humanitarian relief and development. They reflect the changing nature of humanitarian relief where actors are having to deal with longer crises and trying to understand how this impacts delivery models and coherence between immediate support and longer term development initiatives.

Reading of evaluation reports. The analysis of the reports included the following:

- What was purpose of the evaluation—how is this described in the TOR and the report itself?
- What evaluation approach and methodology were used, and how does this approach reflect the requirements set out in the TOR?
- Does the evaluation approach and understanding of accountability in the reports conform to policy statements by the INGOs?
- Did the report include a description of how the findings of the report was disseminated and used, and, if so, how was the information disseminated and used?
- What descriptive terms are used in referring to accountability? For example, does the report recognize the evaluation itself can be used for accountability to affected communities, or is reference to accountability more limited to how accountable the organization was during project implementation?
- How are issues of power represented in the text? In the review both overt and hidden manifestations of power will be considered. As an example, overt language may include a clear statement the evaluation is aimed at more powerful stakeholders such as donors. Hidden manifestations may come from an

interpretation of the type of language or metaphors used, reflecting an evaluation structure set up to respond to the more powerful stakeholders. Examples of this could include evaluation reports rely heavily on the language of NPM.

Interviews

Ten interview participants were contacted from individuals, either external evaluators or NGO staff, named in the evaluation reports. Three additional individuals were contacted at the suggestion of interview participants. Out of the 13 individuals contacted, 11 responded and 10 agreed to interviews. The interviews were all conducted over skype and recorded to allow for detailed review later. Interviews ranged from 35 minutes to 1 hour. The interviews were semi-structured which allowed key issues and themes identified from the evaluation report analysis to be addressed, while ensuring the interesting information from the participants could be followed up on. Interviews were planned specifically for each evaluation report and interview scripts adapted during the ongoing process to reflect themes that emerged from past interviews. The interviews were important for two main reasons. Many evaluation reports did not include information on dissemination and use. To identify how the evaluation was used as an accountability mechanism, particularly towards affected communities, it was necessary to speak with individuals who participated in the evaluation. In addition, a critical element of the interviews of evaluation participants was to try to understand the context in which the evaluation was completed, the approaches to setting the scope of the evaluation, and what the true purpose of the evaluation was. These interviews contributed a greater understanding of the author intentions and the conditions and power dynamics involved in the evaluation.

Although the interviews were developed for individual evaluations, common questions included:

- What did you believe the main purpose of the evaluation was?
- How were the scope of the evaluation, the TOR, and the evaluation questions decided upon? Were stakeholders outside of the NGO included in the process?
- Do you think the field offices/beneficiaries/local stakeholders felt ownership of the evaluation?
- How was the evaluation report used and the findings disseminated to stakeholders?
- Who did the evaluation provide accountability to?
- How do you define accountability?
- Who do you think benefitted from the evaluation?

Data Analysis

The evaluations reports, accompanying documents such as the TORs, and interview transcripts were analyzed with the use of the qualitative analysis software Atlas Ti. The data was analyzed by deep reading and re-reading of the texts to gain an understanding of the meaning and importance of individual elements of data, and situating them within the broader contextual environment of the accountability movement within humanitarian relief. Moving through the hermeneutic circle allowed me to merge new understanding from the text with past understanding from previous experiences and stages of the current research. This ensured new questions were identified as the research progressed. This knowledge fed into the interviews as

they progressed. This allowed me to follow up in later interviews on ideas that occurred in the early interviews. This approach to the data analysis began through the literature review. The literature review allowed me to use past experience with INGOs, and knowledge of current practices, and fuse this with new knowledge gained through research of academic and grey literature. With this approach, I identified important current practices and policies that informed the development of the research questions. Once the research questions were developed, a further review of literature allowed for new understandings of specific aspects of accountability, evaluation, and EHA and a broader understanding of the context of the whole. This process continued using the hermeneutic circle to analyze the evaluation reports and then interview transcripts. The evaluation reports were analyzed prior to conducting the interviews. This analysis contributed to the finalization of the interview questions. After conducting the interviews and analyzing the transcripts, there was a re-reading of the evaluation reports to bring knowledge gained from the interviews into understanding of the reports. The building of knowledge at every stage of the research allowed new perspectives on previous experiences and contribute to a greater understanding of the use of evaluation as an accountability mechanism.

A key element of understanding how evaluations are responding to policy discussions was to consider the evaluation and accountability policies and statements of the INGOs and compare these to the actual practice of evaluation. The websites of the INGOs included in the sample were searched for details on the INGOs approach to evaluation. Evidence included web-pages laying out the INGOs understanding of accountability, INGO evaluation policies and manuals, and policy statements and documents concerning accountability. These contributed to the knowledge gained from the analysis of the reports and the interviews to allow a broader understanding of evaluation practice.

The literature review identified the power imbalance between stakeholders is a key concern within the humanitarian sector. As such I chose a critical hermeneutic approach that recognizes concerns of power and ideology within texts. The data analysis process therefore interpreted how issues of power and marginalization were described in evaluation reports and TORs. This contributed to the understanding of current evaluation practices by grounding the research within the structures of power that exist. The awareness of these issues added a layer of understanding at each stage of data analysis within the hermeneutic circle.

Chapter 4: Results

Despite INGOs policy statements, I found current evaluation practice of EHA does not support the use of evaluation for accountability to affected populations. Communities have limited access to the results of the evaluations and no involvement in planning the scope or questions of the evaluations. Communities are consulted at the data collection stage of the evaluation for their opinions on the intervention and downward accountability in project implementation is included as a criterion.

Evaluation practices focus on utilization and methods oriented evaluation theories. The research found only very limited use of valuing approaches, that is, evaluations that prioritized community empowerment or democratic involvement. This finding suggests INGOs are benefitting the most from current practice along with partner civil society groups. The benefit to affected communities was much more indirect and dependent upon the INGOs identifying accurate findings, learning lessons and implementing changes as a result.

The key findings from the data analysis were:

- Policy and guideline frameworks emphasizing the importance of downward accountability and participation exist.
- The use of evaluation as a direct mechanism for accountability to affected populations is very limited. No involvement of beneficiaries in evaluation design and deciding what questions to ask was found. There was also very limited involvement in beneficiaries in analyzing the findings or even being presented with the results.
- Involvement of local partner civil society groups in evaluations provided the most evidence of downward accountability.

- Evaluations are being used to assess how downwardly accountable INGOs are being during project implementation.
- Beneficiaries are being consulted for their opinions on the project at the data collection stage.
- INGOs appear to be prioritizing use through internal lesson learning.

Evaluation Policies and Accountability Approaches

A review of INGO's policies provided an understanding of the agencies' approach to evaluation and accountability, and illustrates whether INGO evaluations follow their own policies and guidelines.

Evaluation Policies

Of the 8 INGOs included in the sample, 4 INGOs (CARE International, Oxfam, ACF and NRC) have publicly available evaluation policies. Save the Children has a detailed evaluation handbook available which includes extracts from its Management Operating Standards, but does not make public its evaluation policy. Christian Aid and DRC's evaluation policy is not publically available, and Support to Life does not have an evaluation policy¹.

All 5 published documents refer to accountability. Oxfam's policy most strongly indicates accountability to the communities they work in is a priority focus.

“The objective of this policy is to help institutionalize this practice throughout the Oxfam confederation, so that evaluation consistently:

¹ By the standards of all the other sampled NGOs, Support to Life is a very young organization and it is perhaps not surprising they do not yet have a written evaluation policy

- enhances mutual accountability and learning between the communities and partners with whom we work, ourselves and our donors;
- enhances the ability of those people whom we seek to benefit to create opportunities and means to hold us to account” (Oxfam, 2010, p. 1)

“As a rights-based organization, accountability, particularly to the communities we seek to serve, is of the highest importance to us. For Oxfam, accountability requires Oxfam to regularly and honestly assess the quality of its work, share and learn from its findings with primary stakeholders, and apply that learning in future work.” (Oxfam, 2010, p. 2)

Oxfam’s policy suggests evaluations should be an opportunity for communities to hold them to account, and also identifies accountability for Oxfam requires learning from evaluation findings *with* communities. This suggests an active form of accountability where communities are involved in the learning process rather than just being presented with results. The policy also identifies the involvement of communities in setting evaluative agendas. “In addition, Oxfam staff should be open and responsive to emerging opportunities and requests for evaluative exercises, particularly requests from the organizations and communities with which we collaborate.”(Oxfam, 2010, p. 2)

There is a sentence that demonstrates donor priorities may be a significant driver of evaluation decisions: “When making decisions about evaluation priorities, managers should consider... demands for accountability from stakeholders, including back donor requirements in direct financing and co-financing arrangements.” (Oxfam, 2010, p. 2)

In the list of considerations for evaluation priorities, community requests are not mentioned. Despite this, the policy clearly identifies Oxfam’s framing of accountability focuses

heavily on the communities that Oxfam work in, and that Oxfam considers evaluation a mechanism to provide downward accountability.

CARE International's policy and Save the Children's evaluation manual contain less overt references to downward accountability and focus more on the importance of participation in evaluations. CARE International's policy refers to accountability five times in the policy and only once gives an indication accountability has a directional element when on page 4 the policy acknowledges past evaluations had focused on donor priorities, and CARE acknowledged the need for a more holistic approach. Save the Children's manual does identify that, "Evaluations can help us to...ensure accountability and transparency to our stakeholders, including children and carers" (O'Neill, 2012, p. 5). However the manual focuses more on how broad participation of a variety of stakeholders improves accountability, rather than focusing explicitly on beneficiaries and affected communities.

"Getting a wide range of stakeholders to participate in the evaluation builds ownership of the process and promotes accountability." (O'Neill, 2012, p. 9)

"You need to identify all those who have been involved or affected by the project or programme – your stakeholders. You must ensure that children and young people from different backgrounds have the chance to participate meaningfully in the evaluation process" (O'Neill, 2012, p. 28).

Participation is an important element of the evaluation policies of CARE International, Oxfam, and NRC, and the Save the Children manual. However with the exception of the above statement from the Save the Children manual, all of the other policies leave a caveat that acknowledges full participation will not always be possible:

“It is integral to a rights-based approach that participants in the project being evaluated should, whenever and **as much as possible**, be actively included in the planning for, implementation, analysis, reporting and utilization of evaluations.” (CARE International, 2008, p. 5) (emphasis added)

“NRC’s stakeholders, including refugees and internally displaced persons, should **when appropriate** be consulted regarding the identification, planning, implementation and utilization of evaluation projects.” (Norwegian Refugee Council, 2005, p. 8) (emphasis added).

Only Oxfam’s policy and Save the Children’s manual directly link participation in the evaluation and downward accountability. CARE International links participation to a rights based approach, which can be linked to the stakeholder version of downward accountability (Lloyd, 2005) but does not specifically link participation and accountability. A word search in NRC’s evaluation policy reveals accountability is mentioned five times, three of which come from invoking ALNAP’s full name. The introduction does identify one goal of evaluation is to enhance accountability, but the policy does not identify how accountability is defined: “NRC is with this evaluation policy document introducing an evaluation policy that is intended to make further contribution towards NRC’s capacity for organisational learning, performance review and accountability.” (Norwegian Refugee Council, 2005, p. 2)

NRC also has a handbook published later than the evaluation policy and this does detail accountability more clearly than its policy does.

“If the evaluations of humanitarian action are to serve as a basis for accountability, they must:

- Be clear about the organization’s responsibilities, standards and benchmarks

- Clarify what was under the organization's control and what was not
- Focus not only on activities, outputs and outcomes but also impacts
- Express the view of key stakeholders, especially the intended beneficiaries
- Be disseminated more widely, including in the area or operation
- Lead to some form of reward or reprimand" (Norwegian Refugee Council, 2008, p. 36)

Within this statement there are clear indications downward accountability is an important focus. NRC's handbook suggests the views of intended beneficiaries and the dissemination of the results are important requirements for accountability. The handbook also reinforces the statement in the policy that broad participation is important, listing it as a guiding ethical rule: "Broad participation-all interested parties should be involved wherever relevant/possible" (Norwegian Refugee Council, 2008, p. 33). The notion of reward or reprimand is also a very strong statement in the handbook, implying accountability is more than just a moral or voluntary action, but it does not clarify who does the rewarding or reprimanding.

Although analyzing the handbooks of Save the Children and NRC is useful for understanding some of the considerations these organizations believe should go into an evaluation, they are handbooks rather than policies. Handbooks focus on best practices as opposed to policies focusing more on minimum acceptable standards. Comparing handbooks to the policies of other organizations is comparing similar information, but in forms bearing different degrees and kinds of authority.

Christian Aid's access to information policy states their evaluation policy is available on request, but my e-mail did not receive a reply. However, Christian Aid's detailed policy

statements on accountability give an indication of how Christian Aid envisions evaluations contributing to accountability:

“Ensuring that communities are able to influence decision making by enabling affected women, girls, men and boys and other stakeholders to participate in different stages of the project, including

- identifying what needs to change
- designing and implementing projects
- monitoring what is delivered
- assessing impact” (Christian Aid, 2012, p. 2)

The same document requires summaries of evaluations be made available to the communities in which they work. So although their evaluation policy is not published, from this document is it possible to infer Christian Aid’s policy is to support participation and involvement of affected communities in evaluations.

ACF’s evaluation policy is the only policy that does not mention the participation of beneficiaries. ACF explicitly states the direct intended users of its evaluations are its field and headquarter teams. Donors are mentioned as indirect users. There is no reference at all to beneficiaries being users of the evaluation, and indeed no reference throughout the policy to the participation of beneficiaries at any level of the evaluation. ACF does address accountability in its policy: “The ACF EPG is designed to ensure that whilst direct users have a greater say on the scope of individual evaluations, each evaluation directly contributes to the wider accountability and learning efforts of the organisation.”(ACF International, 2011, p. 8)

The policy does not clearly specify how evaluations provide accountability but focuses on the DAC criteria as being the means by which an evaluation provides accountability. The section called “Accountability: the DAC criteria” explains what the DAC criteria are and concludes:

“Evaluators will be expected to produce a rating of the overall programme/project using the DAC criteria. Rating each evaluated programme will serve a dual purpose: 1) enable programmes to monitor the progress made between evaluations (e.g. from one year to the next), and; 2) enable ACF to monitor the collective progress made across the organisation.”(ACF International, 2011, p. 9)

This demonstrates ACF’s belief that the purpose of evaluation is mainly for internal accountability, organization monitoring and lesson learning.

Aside from ACF, those INGOs that make their evaluation policies public emphasize either (and in some cases both) how evaluations can provide accountability to affected populations or the importance of participation at all stages of the evaluation. The locating of accountability within the evaluation policies suggests INGOs recognize the possibilities for evaluation to be a mechanism for downward accountability. The policies also show INGOs make statements that support processes involving communities at all stages of the evaluation. However, in all of the actual policies (as opposed to handbooks) there is a statement indicating participation will not always happen.

Accountability Statements

All of the NGO reports included in the study sample refer to accountability on their websites. Given the many versions of accountability identified in the literature review, it is not surprising the NGOs have both different definitions of accountability. CARE and Christian Aid

have detailed accountability policies published on their websites, while others, in particular NRC, ACF and Support to Life, have much less detail about their views of accountability. Table 1 summarizes the statements made on INGO websites and in policy documents.

Seven out of the 8 INGOs included in the study sample refer to beneficiaries or affected communities in their accountability statements. Only Oxfam and CARE International identify accountability includes being held to account by affected populations as well as holding themselves to account. Save the Children make a similar statement in their accountability handbook, but not their accountability statement on their website:

“At Save the Children we think that real accountability to children and communities involves giving them not only a voice, but also the opportunity to influence relevant decisions affecting whether and how we work with them. And it involves giving children and communities the power to hold us to account in ways that influence the organisation’s policies, priorities, and actions at local, national and global levels. Such influence balances the power that donors and governments (as regulators) have to influence us.” (Munyas Ghadially, 2013, p. 1)

Save the Children is one of four INGOs to identify accountability to the people they work with as a means of redressing power imbalances in the humanitarian sector. CARE International, Oxfam and Christian Aid also refer to power in their accountability statements and policies. CARE International’s most closely aligns with HAP’s definition through specifying a particular definition for humanitarian accountability: “Humanitarian accountability is an appropriate shift of the balance of power back towards disaster affected people.” (CARE International, 2010, p. 1)

Two of the NGOs in particular identify transparency as being a key part of accountability. As with its evaluation policy, ACF’s accountability statement does not indicate

beneficiaries or communities are at the heart of its thinking on accountability. In fact ACF's statement is unique in the sample in that the first half is focused on requiring accountability or action from others, "we directly oversee the implementation of our programs, requiring full access to the communities we assist." The second half focuses on the disclosure of financial information. Although not explicitly stated, the traditional disclosure of financial information in the humanitarian sector has been to donors, and in less detail to those that can access NGO websites and read English or French. NRC's statement of accountability places more focus on providing information to IDPs and refugees:

"NRC is committed to being accountable to our staff and beneficiaries. Transparency is a prerequisite for accountability, and NRC is always aiming at making information about the work we do more accessible and visible. We are working with internally displaced persons and refugees, and place particular emphasis upon the importance of sharing information with them as well as the general public. The information we publish and how we respond to requests for information are important aspects of accountability."

(Norwegian Refugee Council, 2014)

Access to information is certainly a key basic requirement for accountability. NRC does not indicate how it makes information available and the statement is on a web-page focused much more on the disclosure of financial information through the International Aid Transparency Initiative (IATI). The IATI is an initiative to improve the disclosure of comparable and timely financial information by international aid organizations. Although this is clearly important for some levels of accountability, it does little to increase accountability to affected populations who in many cases have very limited internet access and cannot read English. The same is true for publishing evaluation reports. This is important for horizontal accountability and encouraging

lesson learning and the sharing resources, but does not improve accountability to affected populations. The reports are generally in English, French or Spanish, and not the local language of the affected community.

Most of the NGOs have signed up to many of the different accountability initiatives. Table 1 shows the most common accountability initiatives NGOs have signed on to. HAP and People in Aid have an additional certification process NGOs can choose to go through, and those certified are also indicated in table 1. The table also summarizes the accountability statements available on websites and in publically available documents. As can be seen, there are wide variations in defining and describing their approaches to accountability between the NGOs.

All but one of the organizations are members of HAP. The HAP benchmarks are particularly relevant as they provide an understanding of the nexus between downward accountability and evaluation, highlighting standards for downward accountability that are specific to evaluation. The main HAP benchmarks relevant to evaluation are sharing of information, participation, and learning and continual improvement. Specifically, benchmark 3.2 is: “The organization shall share with the people it aims to assist and other stakeholders information appropriate to their needs, including: ... its goals and project objectives, expected results, with the time frame, and a financial summary, as well as summaries of evaluations and progress reports” (2010a, p. 35). Benchmark 3.3 is “The organization shall ensure that information specified in 3.2 is presented in languages, formats, and media that are appropriate for, accessible to, and can be understood by the people it aims to assist” (2010a, p. 35).

NGO	IFRC Code of Conduct	HAP Member	HAP Certified	ALNAP Member	People in Aid Verified	People in Aid Member	INGO Accountability Charter	Accountability Statements
CARE	Yes	Yes	No	Yes	Yes	No	Yes	We define accountability as the means by which we fulfill our responsibilities to our stakeholders, such as donors and our beneficiaries, and the ways in which they may hold us to account for our decisions, actions and impacts
Oxfam	Yes	Yes	No	Yes	Yes	No	Yes	We believe that non-governmental organizations should be accountable to the communities in which they work, to partner organizations, and to those from whom they receive support.
Save the Children	Yes	Yes	No	Yes	Yes	Yes	No	We take personal responsibility for using our resources efficiently, achieving measurable results, and being accountable to supporters, partners and, most of all, children.
Support to Life	No	No	No	No	No	No	No	STL is accountable to its stakeholders in projects including individuals, communities, partner organizations, and governments.
NRC	Yes	Yes	No	Yes	Yes	No	No	NRC is committed to being accountable to our staff and beneficiaries. Transparency is a prerequisite for accountability, and NRC is always aiming at making information about the work we do more accessible and visible.

NGO	IFRC Code of Conduct	HAP Member	HAP Certified	ALNAP Member	People in Aid Verified	People in Aid Member	INGO Accountability Charter	Accountability Statements
DRC	Yes	Yes	Yes	Yes	Yes	No	No	In the DRC, we wish to be accountable not just to our donors and membership, but also –especially – to those we try to help. We have many instruments for this, like participation, feedback, complaints-handling and reporting. A key component, however, is informing our beneficiaries and other stakeholders about our Accountability Framework. It defines what they should hold us accountable for.
ACF	Yes	Yes	No	Yes	Yes	No	No	We directly oversee the implementation of our programs, requiring full access to the communities we assist. We are committed to a policy of transparency and disclosure by ensuring that key financial information is publicly available and that our programs undergo external impact evaluations.
Christian Aid	Yes	Yes	Yes	Yes	Yes	No	No	It means holding ourselves openly responsible, in ways that involve our key stakeholders, for what we believe, what we do and say we will do – and for showing what we have done compared to what we said we would do.

Table 1. Summary of INGO membership/signatures on global accountability initiatives.

Benchmark 4.2 is “The organization shall develop and put in place processes appropriate to the context so that the people it aims to assist and other crisis affected people provide feedback and influence 1. initial assessment; 2. project design, deliverables, criteria for selecting target groups and the selection process; 3. project implementation; and 4. monitoring and evaluation” (2010a, p. 39). Finally, Benchmark 6’s introduction states “continual improvement is achieved through an effective monitoring and evaluation system, which ensures regular reviews of the work, impact and effectiveness of the organisation, and that identifies lessons for improving future operations” (2010a, p. 48), with 6.3 being relevant to the findings from this research: “The organisation shall include in the scope of evaluations an objective to assess progress in delivering its accountability framework.” (2010a, p. 49).

In policy terms there is both a commitment to downward accountability and a belief evaluation can contribute to providing this form of accountability. Accountability statements often recognize beneficiaries as being the most important group that should receive accountability, accountability initiatives offer a framework for how evaluation fits into downward accountability, and evaluation policies highlight the importance of the involvement of beneficiaries and communities in the design, conducting and analyzing of evaluations.

Affected Community Involvement in the Evaluation

The research looked at community involvement in the evaluation at several different stages. These included planning the evaluation through contributing to identifying the scope, designing the questions and giving input into the TOR, giving data to answer questions, collecting data, analyzing data and receiving the results. Table 2 summarizes the involvement of beneficiaries and affected community members in each of these stages. These are described in more detail below but in summary the data show community involvement is low in most stages

of the evaluation with the exception contributing data to the evaluation through focus groups, surveys and participant interviews. The only exception to this lack of participation at other stages of the evaluation is through the involvement of local civil society partners such as NGOs and CBOs.

Terms of Reference and Evaluation Planning

In the review of NGO evaluation policies and accountability policies affected community involvement in planning, design, data collection and analysis of data is an important but not compulsory element of an evaluation and one might expect this emphasis to be reflected in TORs. However the review of the TORs did not include consulting the affected community or beneficiaries on the design of the evaluation. Evaluators and NGO staff interviews provided data on the nature of community involvement in the evaluation but none identified the inclusion of affected communities in developing the TOR, deciding the scope of the evaluation, or designing the evaluation questions. In general, these evaluations were planned and conducted with limited input from stakeholders outside of the organization and no input from the affected communities. What limited consultation there was, was with external stakeholders specifically donors and partners. Evaluators and NGO staff indicated the lack of consultation with beneficiaries and community members in developing an evaluation is not limited to the evaluations in this study, but is widespread across the field.

Evaluation	Designing the Evaluation	Consulted during Data Collection	Helping Collect Data	Analyzing Data	Receiving Results
NRC	No	Yes	No	No	No
DRC	No	Yes	No	No	No
STC	No	Yes	No	No	No
Christian Aid	No	Yes	No	Yes-Local NGOs ²	Yes-Local NGOs
ACF	No	Yes	No	No	No
Oxfam-Kenya, Gaza & Haiti	No	Yes	No	No	Yes-Local NGO
Oxfam-South Sudan	No	No	No	No	No
CARE-Sahel	No	Yes	No	No	No
CARE-Gaza	No	Yes	Yes-CBOs	Yes-CBOs	Yes-CBOs
Support To Life	No	Yes	No	No	No

Table 2. How beneficiaries and affected community members are involved in the evaluation process.

² Local civil society involvement is indicated where qualifier is put after ‘Yes’

Question: “generally on your evaluations do you ever really see evaluation where the communities give input into the questions and design?”

Answer: “No, almost never... I've once been asked to prepare a report for, to feedback to the community who participated...so on one or two occasions, I would say I've been involved at either end, normally it's feeding back on an evaluation to communities, but I'm not aware of an evaluation I've been involved in, I can't think of one, where the community has had input into the questions or into the TOR, no.”

Additionally, evaluators believed most TORs were based on generic organizational templates with limited attention to the individual project. There is no evidence in any of the TORs, evaluation reports or from discussions with the evaluators or NGO staff that affected communities were involved in giving input into the development of the evaluation TORs.

“it's a fairly stock standard TOR, it's just asking the standard DAC criteria questions in some way.”

“Question: Do you know whether other stakeholders were involved in the TOR?

Answer: That's a good question, so it depends I think on how we start to talk about being involved. So if you took it from a strong gender perspective or you took it from a kind of empowerment evaluation perspective then they probably weren't, but if you took it from a kind of the reality of the way that evaluation is practiced, then I think they had spoken to their government counterparts, and they had an implementing partner there that I think they had spoken to on a kind of courtesy type consultation but the evaluation questions were shaped by (the NGO), which is the way that most of these things happen, so in that sense it wasn't participatory.”

In most cases the evaluation criteria were set by the NGO and included in the original TOR, but there are two cases where the evaluator gave input into the criteria. The NRC evaluation report methodology section describes how the evaluator persuaded the commissioning manager to expand the criteria of the evaluation to be more in line with NRC's evaluation policy. Similarly the CARE Sahel evaluation report highlights the exclusion of connectedness as a criterion from the initial TOR but inclusion in the evaluation itself. However there is no evidence for any of the evaluations either in the reports or from the interviews that affected community involvement influenced the evaluation criteria.

The process for developing evaluation questions varied. In some cases evaluation questions were set within the TOR and the evaluator's role began at the stage of designing indicators and tools to answer those questions. In other evaluations, general overarching questions were set within the main criteria in a TOR but the specific evaluation questions were developed by the evaluator, usually as part of an evaluation matrix or inception report. Many TORs and supporting documents annexed to the reports include timetables for the evaluation. These show evaluation questions and the design of indicators and tools are developed before the evaluators come to the field during the "desk review" stage before the evaluator was in the country where the program was being implemented. For example, ACF's TOR requires: "A full methodology including approach, stakeholder analysis, evaluation matrix, FGD and KII questionnaires, a detailed work plan and interview list will be provided in an Inception Report to the Evaluation Team and Target Users during the preparatory stage." The annex detailed the actual itinerary for the evaluation showing the inception report is provided before the evaluator was in the country, thus making it impossible for the evaluator to use participatory methods to design the evaluation. This practice was confirmed by interview participants, none of whom

identified the involvement of communities in developing evaluation questions or tools. Therefore, as with the development of the TORs, the development of evaluation questions did not involve project beneficiaries, a point highlighted by this evaluator: “No actually, I think that [is] one of the missing parts, one of the important missing parts regarding the development, was the lack of participation from the right holders side... I mean the participation was missing I would say.”

The use of evaluation steering committees or guidance groups is one means of ensuring community members can participate more deeply in the evaluative process. For example, the Save the Children evaluation manual identifies one method for ensuring children’s participation is for them to be members of an advisory or reference group (O’Neill, 2012). However there was no evidence NGOs are forming steering committees or guidance groups that include members of the affected community. Only 3 of the evaluations refer to a steering committee in the reports, and all of those consist only of NGO staff members.

The contrast between evaluation policies and accountability commitments and actual practice at the front end of an evaluation is marked. Despite the stated policy of many NGOs to include beneficiaries and affected communities in the design of evaluations, this was not the case in any of the sampled evaluations. While it is possible the NGOs would argue the design is based on feedback gathered during the projects as part of a monitoring process, this is not described in the reports. This raises the question as to whether evaluations are answering or investigating the questions that the affected communities want addressed in the first place.

Collecting Evaluation Data

The collection of data came from a variety of sources and in a variety of forms: monitoring numbers; donor reports and complaints; interviews and focus groups with

stakeholders including project beneficiaries, community leaders, and government officials; surveys; and formal and informal discussions with staff members. The significant feature of this stage of the evaluation is that project beneficiaries, and often other community members, participated in the giving data to the evaluators. Only one evaluation, Oxfam's desk based evaluation using its Global Humanitarian Indicator Tool did not obtain data from project beneficiaries. For some others, data collection from beneficiaries and community members was limited by factors exogenous to the evaluation. Because of security concerns CARE's evaluation in the Sahel only consulted beneficiaries in Chad and not in Mali and Niger, and Christian Aid's data collection process was limited to only certain accessible villages. Despite these limitations, it does seem clear that beneficiaries and to a lesser extent other community members as sources of data is considered important by NGOs and this is one area where NGO policies are matched by actual evaluation practice.

Table 3 summarizes how the reports collected and presented data. Of the 9 that collected data from beneficiaries, 2 did not specify the number of participants they spoke to although both of these reports explained how they identified beneficiaries to talk to. Most disaggregated data by at least gender, with fewer identifying how many direct beneficiaries and how many other community members they spoke to, and three disaggregate by adult and child. Seven gave at least a brief explanation of the sampling process, although the level of detail varied widely. For example, the description in the Christian Aid evaluation is fairly brief:

“Field visits to beneficiary villages. These were selected by CA and partners and confirmed with the evaluation team. Primary reasons for selection concerned accessibility, both physically (some locations were remote) and security-wise (i.e. where

it was safe to go). The team visited at least one project of all partners.” (Morgan, Naz, & Sanderson, 2013, p. 8)

NRC’s report includes a four page explanation of the sampling process.

The means of collecting data directly from beneficiaries and other community members included focus groups, interviews and surveys, with focus groups being the most common. Seven of the nine evaluations that collected data from beneficiaries and other community members used at least one focus group. ACF did not organize formal focus groups but had informal discussions during site observation visits, which was in recognition of the difficulties in asking community members to attend a formal meeting only one month after the typhoon had struck the Philippines. CARE’s evaluation in Gaza had group interviews with staff but the discussions with beneficiaries came through the collection of individual stories of change.

Although there was some involvement from local community members in helping to identify beneficiaries to interview, only CARE’s evaluation in Gaza involved community members in the actual collection of data. In this example, CBO members were involved in collecting the stories of change. There was some evidence of community members being used as a resource for identifying people to talk to. For example, NRC’s evaluation team asked headmasters to identify participants for individual interviews and Christian Aid asked partners to help select villages and project sites to visit. However, with the exception of CARE’s evaluation the involvement of communities in collecting data for the evaluations was limited to supporting the identification and finding of participants.

Evaluation	Collects Data from Beneficiaries	Gives numbers of participants consulted	Disaggregates by:			Explains participant selection process
			Gender	Beneficiary and other community	Adult and child	
NRC	X	X	X	X	X	X
DRC	X	X ³	X			X
STC	X	X	X	X	X	X
Christian Aid	X	X	X			X
ACF	X					X
Oxfam-Kenya, Gaza & Haiti	X					X
Oxfam-South Sudan						
CARE-Sahel	X	X	X	X		
CARE-Gaza	X	X	X ⁴	X	X	X
Support To Life	X	X	X ⁵	X		

Table 3: Details of beneficiary and community consultation in the data collection stage of evaluations.

³ DRC's report gives the number of beneficiaries that participated in a survey but does not give numbers of focus group participants.

⁴ CARE's Gaza report does not tabulate the split of male/female and adult/child beneficiaries consulted, but does list this in the heading of each story of change included in the annex. Therefore it would be possible to get disaggregated numbers if one were to go through the annex and tabulate this.

⁵ An annex to the Support to Life report disaggregates the beneficiary participants by gender but not the other community members, such as shop owners.

Negotiating and Presenting the Results

The negotiation and presentation of the results was mostly conducted with staff members of the commissioning NGO. The evaluations where the results were negotiated with members of the affected communities were projects that partnered with local civil society groups. In both CARE's Eye to the Future project in Gaza and Christian Aid's Pakistan Flood Relief program, local project partners were involved in feedback workshops where results were presented, analyzed and agreed upon. In the case of Christian Aid, the partners participated in a meeting on the last day of the evaluation and recommendations were negotiated. The CARE evaluation in Gaza used the Most Significant Change approach where representatives from CBOs worked with CARE staff to read, discuss and select the stories they believed best represented the change the project had brought. Even in these two evaluations the involvement of partner organizations did not extend to including beneficiaries. Given local CBOs are situated within the target communities, the involvement of these organizations brings the results closer to the beneficiary communities, but still does not involve beneficiaries directly in negotiating the results.

In the other evaluations there was no evidence the evaluation itself involved the negotiation or analysis of the results with the beneficiaries, although some of the evaluators did acknowledge the possibility the NGO had undertaken this process afterwards. None particularly expected this to have taken place though, beyond possibly the presenting of findings in some form to the community.

When asked to reflect on evaluations in general in a humanitarian context, evaluators believed very few NGOs had negotiated the results with the affected communities. In fact, there was a general belief that results were rarely even presented to the community.

“It's not impossible, I've done it with some organizations before but my general feeling is it is an added task which if you are clear about that that is the purpose then you generally do it but if it's not written on the instructions, it tends not to be done. So that's my perception, that I doubt very much that the results of this evaluation are fed back in any way to communities, it's possible that I might be doing people an injustice but that would be my guess.”

Two evaluators also highlighted the fact reports were generally not translated into the language of the beneficiaries. The HAP benchmark on information sharing requires information be shared in languages and formats affected populations can access and summaries of evaluations should be provided. The INGO statements and policies on accountability and evaluation also make commitments to presenting information to affected communities. For example, NRC's accountability statement refers to how NRC is working to make information more accessible and visible for beneficiaries (Norwegian Refugee Council, 2014), and Christian Aid's accountability framework requires summaries of evaluation reports are made available to poor and marginalized communities (Christian Aid, 2012). However, the belief of interviewees was this was not taking place in these evaluations, and rarely takes place at all within the sector.

If evaluators are not presenting results to the local community are they involved in presenting the results to the other stakeholders? Eight of the TORs included a requirement to give a debriefing in the field prior to the end of the consultancy, and a further three also specified a presentation be given to headquarters staff at the end of the evaluation. Only five of the reports provide details about how the results were presented to stakeholders at the end of the evaluation, but interviews with the evaluators and NGO staff provided more information. A common strategy was a debriefing on the last day of the evaluation with staff at the country office that

may or may not include the field staff involved in the program implementation. Many evaluators also alluded to informal debriefings with field staff as the evaluation developed. CARE's evaluation in Sahel is the only evaluation that describes a large scale learning workshop that included senior management from the three evaluated countries, the emergency department, and various European and North American CARE funding offices.

The inclusion of other stakeholders is much more limited, with the exception of direct implementing partners such as CBOs and local NGOs. Three of the TORs (DRC's, NRC's and Support to Life's) refer to the inclusion of other stakeholders in a debriefing session. The language used suggests beneficiaries were not the stakeholders in mind when the TOR was written. For example, DRC's TOR requires the evaluator "present a Draft Evaluation Report (2nd draft report) to DRC, ECHO and other stakeholders in workshop to facilitate sharing of evaluation results with a view to incorporate inputs from project stakeholders in the final draft." The DRC TOR is the only one that explicitly mentions that donors, in this case ECHO, will be included in the presentation of findings. Other TORs simply refer to "key stakeholders" or "internal and external stakeholders." The vague language in the TOR means beneficiaries are not explicitly excluded, but the responses of the interview participants highlighted in this section indicate beneficiaries were not present at the feedback meetings.

From reviewing the reports and interviewing evaluators and NGO staff, it is clear external evaluators' involvement in the evaluation process ends after a debrief session with staff members and integration of feedback from that session into the final report. There was very little evidence evaluators are asked to support the NGO in translating the findings into action. Although all the reports contain recommendations for future programming, in most cases these are left with the NGO to translate into action without the support of the evaluator.

Evaluation Criteria

The DAC/OECD criteria were the most common set of criteria used in the evaluations. In the sample of evaluations the 5 development criteria and the 7 emergency response criteria were used, although a number of evaluations used additional criteria. The exception to this is the Oxfam evaluations. Both the South Sudan evaluation using the Global Humanitarian Indicator Tool and the evaluation of Emergency Food Security and Livelihoods program relied very little on the DAC criteria. In both Oxfam's evaluations, timeliness (a key element of appropriateness) was an additional criterion; in the Food Security and Livelihoods evaluation impact features were an added criterion; and in the South Sudan evaluation coverage is an additional criterion.

Table 4 summarizes the criteria used in each of the evaluations. To be included, the criteria had to be explicitly included in the evaluation reports.

Evaluators and NGO staff identified the DAC criteria as providing a structure for evaluations that supported comparison among evaluations and gave a common language or standard for evaluations to be judged by:

“I think it is useful, it gives us a common language and a common frame of reference”
“the idea of the DAC criteria is to give some standards to evaluations and that's why people just them.”

Those that considered the DAC framework as supporting accountability and believed the accountability the DAC criteria provided was an upward accountability to donors: “I think the accountability to the donors is definitely through the DAC criteria because that's just the donor's way of comparing one operation to another let's say. And that's why the DAC criteria are there.”

Evaluation	Relevance/ Appropriateness	Connected- ness	Coherence	Impact	Efficiency	Effectiveness	Coverage	Sustainability	Accountability	Other
NRC	X				X	X		X		X
DRC	X				X	X				X
STC	X	X		X	X	X		X		
CA	X			X	X	X			X	X
ACF	X					X			X ^c	X
Oxfam- Urban	X			X					X	X
Oxfam- S.Sudan	a	b					X		X ^d	X
CARE- Sahel	X	X	X	X	X	X	X			
CARE- Gaza	X			X	X	X		X		
STL	X			X	X	X		X		

^a Oxfam's GHIT includes timeliness which is a part of ALNAP's definition of appropriateness but does not feature the other elements of appropriateness.

^b Oxfam's GHIT includes a focus on "one-program" approach which seeks to assess the inter-connectedness of Oxfam's humanitarian response and development programs.

^c ACF's criterion of beneficiary participation is considered synonymous with downward accountability based on the definition given within the criterion.

^d CARE Sahel and Save the Children evaluations included accountability to affected populations as indicators within the more traditional DAC criteria (STC within relevance and CARE within efficiency)

Table 4. Criteria used in evaluations.

However it is interesting to note both the interview participants and a review of the evaluation reports and supporting documents highlights most agencies and evaluators are defining relevance in line with the definition in the emergency response document. This defined relevance as: “Relevance is concerned with assessing whether the project is in line with local needs and priorities (as well as donor policy).” (ALNAP, 2006, p. 22) This definition is more in-line with the notion of downward accountability, as taking responsibility for acting in accordance with the needs and wishes of the target community rather than the original DAC definition of relevance that focuses more on whether the project achieved the goals of the project.

Save the Children’s evaluation specifically identifies this definition in the report and CARE International’s Sahel report instructs readers to refer to the ALNAP document that defines relevance in a humanitarian context. NRC’s evaluation relies on NRC’s definition of relevance: “When measuring relevance one checks whether the objectives, project design and activities are consistent with the humanitarian and protection needs and the situation in general” (Norwegian Refugee Council, 2005, p. 7). This definition focuses on the needs within the local context rather than policy goals of NRC or a donor.

Support to Life’s evaluation does not define relevance but the questions in the TOR situate relevance within community need. Only one of the questions refers to the identification of clear objectives and the others focus on beneficiary need and local context:

“Relevance and Appropriateness

- Were the assessments undertaken appropriate for the identification of real needs?
- Was sufficient attention given to the identification of clear objectives and activities?

- Was the assistance appropriate in relation to the customs and practices of the affected population?
- Was the transfer modality the most appropriate to meet beneficiary needs?
- Was the transfer modality the most relevant to the context (host community interaction, market conditions, availability of food, supply chain, risks, gender consideration, seasonal factors, etc.)?” (Kugu & Oksak, 2013, p. 39)

The section on relevance in the report also focuses on how well the project responded to community needs, how the project adapted to community feedback and how involved the community was through project participation.

The TOR in the DRC evaluation defines relevance as “To what extent has the project intervention conformed to the needs and priorities of target groups, and the policies of DRC and the European Commission?” (HERALD Consultants, 2012, p. 47). However the evaluation itself focused more on identifying whether the project met the preferences of the beneficiaries:

“The relevance and appropriateness of the project (wet feeding project) were assessed based on; beneficiaries’ perception on conformity of the food served at the wet feeding centers to the local food culture, utilization, dependency, dignity and respect” (HERALD Consultants, 2012, p. 15)

“The relevance and appropriateness of the project (cash transfer project) was examined by considering its contribution in debt reduction, cash usage and preference especially when compared with other options.” (HERALD Consultants, 2012, p. 23)

The interpretation of relevance in Christian Aid’s evaluation was closest to the more traditional development definition. Within the criterion of quality and relevance, the evaluation

answered questions about whether the project met the aims of the funding appeal, the quality of the work, how timely the intervention was, the capacity of local partners, and how well gender considerations were mainstreamed into the programming. Christian Aid covered beneficiary needs to some extent in the downward accountability criterion, which includes focusing on how beneficiary needs were assessed and how they were able to alter building designs to fit their particular needs.

The effects of the downward accountability movement can be seen in the inclusion of downward accountability in project implementation as an additional criterion. This partly conforms to the HAP Benchmark 6.3 requiring NGOs to include the assessment of delivering on their accountability framework. Although none of the evaluations systematically compared the program against the accountability framework of the organization, the inclusion of downward accountability as a criterion or indicator within a criterion does demonstrate this concept is considered important in program implementation. Christian Aid and both Oxfam evaluations included downward accountability as a stand-alone criterion. ACF included “the level of beneficiary communication during the response and to what extent beneficiaries and local communities are involved in the design and implementation of ACF operations” as one of the key areas in the scope of the evaluation. This can be interpreted as including many of the elements of downward accountability NGOs include in project implementation. The Save the Children evaluation included the assessment of downward accountability in the project implementation as an evaluation question as part of the relevance criterion, and the Care Sahel evaluation more implicitly included downward accountability in program implementation within the efficiency criterion.

These findings suggest NGOs see evaluation as a means to assess the extent to which implementation teams are ensuring downward accountability mechanisms operate within their projects. By including downward accountability as a standard criterion or question in an evaluation NGOs signal evaluation is seen as a means to mainstream downward accountability within country offices. The actual level of downward accountability in the projects or programs was given mixed reviews in the evaluation reports. Christian Aid's and ACF's broadly judged the projects to have approached downward accountability effectively:

“Overall there was evidence of a good level of men and women ‘being heard… Access to information was high… A complaint and response mechanisms were facilitated by all partners for the affected people to give feedback on the services.” (Morgan et al., 2013, pp. 20&21)

However, many evaluations reports identified weaknesses in downward accountability in the project as well as more generally in the monitoring and evaluation procedures suggesting despite the increased attention not all NGOs have not yet become successful in instilling a cultural of downward accountability within their implementation teams. This is reflected in Save the Children's report:

“While more will be presented on SC’s M&E systems and tools later in this report, it is important to state that there are currently no formal structures or mechanisms in place for SC and target groups to share, reflect upon, analyze, and make joint decisions on program strategies and results. SC staff state that they do share program results with the target groups during visits and meetings, which is good. SC does undertake routine needs and situation assessments, which is also good. However, the consultant did get the impression

that the data collected is very much viewed as “SC’s” by both SC staff and the target groups.” (Karlsen, 2012, p. 9)

On a more specific issue, Oxfam’s evaluation of urban programs in Kenya, Haiti and Gaza noted in Nairobi, “There was a complaints mechanism for the programme, but this was not in practice very formal or elaborate. During the targeting process, the opportunity to complain was not formalised, and was available in practice only through the community health workers (CHWs) and partner staff who had been involved in the targeting. This meant that most people were very unlikely to complain.” (Macauslan & Phelps, 2012, p. 70)

There is evidence NGOs are utilizing evaluation as a mechanism to assess downward accountability, even if that accountability is not being fully provided in the project.

Analysis

The first question to address is who are the evaluations providing accountability to? This obviously comes with a caveat that not all the evaluations provided accountability, although publishing the reports provides a very basic level of accountability. The data suggest there are a number of groups who are the recipients of accountability: the INGO itself, donors, the public in the donor countries, and affected communities.

INGOs

Evaluators and NGO staff most commonly identified the evaluations as providing accountability to the NGO itself. This is something not explicitly stated in the reports but came from the impressions of the evaluators and NGO staff and from interpretations of the language used in the reports. Internal accountability has many forms, i.e. accountability to staff, accountability to an organization’s values and missions, accountability from implementers to

management and vice versa, and from field offices to head offices and vice versa. The accountability identified by the interview participants and in the reports was mainly general organizational accountability to the organization's mission and values, and accountability from field offices to headquarters.

"I think in terms of this evaluation, a major focus was (the NGO) holding itself accountable to its own mission and policies."

"So on paper it's accountability maybe to the organization itself to its own standards"

"Country to HQ, yes probably a far stronger accountability, that's where the key conversations are happening and that's where the management response is obviously monitored"

Donors

There is evidence in the reports and from the interviews NGOs see evaluation as serving an accountability purpose to donors. DRC's report states as a purpose of the evaluation: "The evaluation was intended to provide information and an opportunity for learning and accountability purposes, for the donor but importantly also for the rights holders at community level" (HERALD Consultants, 2012, p. 5). In some reports the concept of accountability to donors can be seen in assessing the extent to which the project met the goals and objectives set out in the project proposal. For example, the Save the Children evaluation's purpose was, "... to assess the degree to which the 'Programme' met the objectives as outlined in the project proposals/log frame, with particular emphasis on appropriateness, timeliness, efficiency and effectiveness of the interventions carried out." (Save the Children, 2012, p. 1) The objective of CARE's Gaza evaluation was "... to assess whether the set targets and anticipated outcomes of E2FII were achieved, and to determine the mitigating factors that may have impacted on these

results." (Shah, 2013, p. v) The CARE report is the only one to include reference to donors when justifying the methodology: "An added benefit of MSC is it provides detailed, specific and clear stories of success that donors increasingly value." (Shah, 2013, p. XXIX (annexes))

Interviewees concurred the evaluations were often considered to have a donor accountability component to them. This seemed to be a side-effect of the evaluation, that the evaluation fulfilled an accountability function to donors in addition to the main purpose of the evaluation. "[The evaluation] was commissioned by the headquarters not by the local office, so the headquarters came at it from the sort of point of view of can we learn something the country that we can learn elsewhere and they probably had to do an evaluation anyway for accountability purposes to their donor I imagine so you know, two birds with one stone." Even if the thrust of the evaluation is not accountability in general or specifically accountability to donors, NGOs recognize the reports have a useful function of demonstrating accountability to donors. The reports can be used to demonstrate to donors that the NGO has implemented what it contracted to do, and how efficiently they were used. NGOs show they are learning from mistakes and difficulties in a program by issuing a management response to the recommendations. So while the main purpose of the evaluation may not have been accountability to donors, the reports still offer an opportunity of demonstrating accountability to donors.

All of the TORs in the sampled evaluations required evaluators produce a written report. In most cases an outline of the report sections is included in the TOR and in some cases there are maximum page requirements. Most of the TORs also ask for a presentation of the results to field and/or headquarter staff, with some also specifying to key stakeholders. Generally the key stakeholders are not clearly identified but the reports would suggest beneficiaries or other community members are not included as key stakeholders to whom the results should be

presented. None of the TORs asks explicitly for as an end product that can be presented to beneficiaries or the community.

When reflecting on previous evaluations one experienced evaluator said he had never or only very rarely been asked to present findings to the community or develop a suitable reporting format. “I remember one evaluation where we prepared comic books for the kids and it got picked up once or twice as being this great step forward. But it obviously isn't a great step forward as apart from that one there's only been one or two that I've ever come across where there has even been feedback given to communities based on the findings so it's still an outlier.” This evaluator used an innovative and experimental approach to report to the affected communities and this was regarded by the NGO as a useful new idea, but it is not one that has been taken forward since by the NGO and there was no evidence of this or other innovative approaches to reporting in the evaluation reports sampled. One of the NGO staff referred to other recent evaluations conducted by the NGO where the evaluator was required to develop a reporting format suitable for the affected community. This demonstrates reporting approaches that are more inclusive and aimed at the affected population do exist, but there is currently a limited use of them.

Public in Donor Countries

Two evaluators perceived the evaluations provided a very weak form of accountability to the public in the country the NGO’s headquarters were based in.

“I guess potentially (accountability) to some supporter because they do make them public documents so we can all peruse them to the extent that we want to or that we're interested to and makes decisions about whether we're going to put our money into the pot when we

next see them outside the shops or wherever, but fairly weak I guess. But, well I say weak but there are not many other agencies who make these documents quite so public.”

The respondent here identifies how the evaluation could be used for accountability purposes, but doubts this was likely and also highlights the problem of unpublished evaluation reports within the sector. Indeed many NGOs do not publish evaluation reports at all which ensures evaluation is not used as a mechanism for any sort of public accountability. Therefore, the NGOs that do publish evaluation reports are engaging in greater transparency in this respect and potentially making evidence available for accountability to taxpayers in the donor country or NGO headquarter country should they chose to use it.

The exception to this may be the potential for watchdog organizations to use the evaluations to provide more manageable information for the public. An example of such a watchdog function is Charity Navigator, which rates non-profits with the goal of helping the American public chose non-profits when they make donations. Charity Navigator currently includes accountability and transparency to the American public as a criterion in its ranking and will expand this in 2015 to include local community voice. Reviewing evaluation reports is part of their methodology and so these reports could have an indirect accountability function for the general public in America (and other countries that have comparable watchdogs) (Charity Navigator, n.d.).

Affected Communities

There was a consensus among the interview participants that the evaluations did not provide accountability to the affected communities. Some saw no accountability to beneficiaries from the evaluation: “I think the idea of being accountable to the beneficiaries is not really there let’s say. It’s not a beneficiary led exercise. You know you speak to some beneficiaries during

the course of the evaluation but mostly you are with staff and some external stakeholders.”

Others identified very limited weak version accountability: “So potentially there is some level of accountability at a sort of community level, although it’s pretty weak considering they … may or may not have anything fed back.”

The exception to finding no accountability to affected communities came from project evaluations of projects implemented in partnership with local organizations such as CBOs that were involved in the evaluation process. The most notable example was CARE’s evaluation in Gaza in which CBOs were actively involved in negotiating the stories of change to be included in the report. Other evaluations such as Christian Aid’s evaluation in Pakistan included elements of involvement by local NGOs, although in many cases these were large national NGOs rather than grass-root CBOs. How much these organizations are representative of the affected communities would require a deeper assessment of their links to the community and local power dynamics than this research can offer. Indeed the inclusion of these organizations poses an interesting paradox by adding an extra organizational layer between the INGO and the intended beneficiaries of their projects while at the same time potentially increasing the participatory involvement of communities in the design, implementation and evaluation of the interventions as well as improving information feedback. How this impacts the use of evaluation as a mechanism for accountability to affected communities and intended beneficiaries may depend on actual selection of the local partners in the first place, and how representative these organizations are of the marginalized and vulnerable in the affected communities. Although this research did not analyze the types of CBOs selected or how accountable they make themselves to their communities, their involvement in the analyzing of evaluation findings illustrates one potential strategy for using evaluation for downward accountability.

Evaluators agreed simply being part of the data collection process in an evaluation did not provide meaningful accountability to beneficiaries. When asked if participating in focus groups or individual interviews about an intervention could enhance accountability to affected communities, a few evaluators indicated this was possible to a very limited extent. Most felt whilst the collection of beneficiary opinion in the evaluation was necessary condition for accountability on its own this strategy did not provide accountability. Participation that is limited to extracting data is not truly participatory, and as such does not provide accountability:

“You know, this is, you're touching upon a point which is a very important point for me professionally is that accountability is not talking to somebody, accountability is engaging somebody in a process, and so this was what I was coming up against throughout my evaluation, and it's not just unique to this NGO, across the board is that when you ask people about participation in the program is that they say, "yeah, yeah, they come to the trainings, yeah, we asked them" and for me that's not participation. That's just extractive, and when only extracting information from this target group, you're not engaging them as decisions makers in their own sort of futures and their future engagement in the program, so no I don't think that by me extracting information from the focus groups that that was in anyway creating any kind of accountability. It was just one more outside coming in and asking them for questions and them not knowing where this information is going so I would say no.”

Evaluation Approaches and Methodologies

In order to determine if evaluations fostering downward accountability used particular specific evaluation approaches and methodologies, the evaluation reports were analyzed using Alkin and Christie’s (2004) evaluation theory tree. This analysis indicated there was not one

common approach to evaluation used. Some of the evaluations could be placed within specific approaches, even if most did not explicitly identify an approach in the report. Some were explicitly identified as democratic evaluation, real time evaluation or utilization focused evaluation. Five of the evaluations had strong elements of more than one approach, but Alkin and Christie's (2004) evaluation theory tree is still a useful tool for characterizing these evaluations. Alkin and Christie observe many evaluation theorists might reside on more than one branch, so too some of the approaches used in the sampled evaluations could fit into multiple branches.

None of the evaluation reports make explicit reference to the approach of a particular evaluation theorist. The methodology sections of the reports included varying degrees of detail. For example, the CARE evaluation in Gaza gives a detailed description of the most significant change approach and its relevance to this evaluation, and NRC's evaluation report outlines process tracing in detail, but did not specify an evaluation theory. Indeed the only reference in the supporting documentation is ACF's referral to Michael Q. Patton's Utilization Focused Evaluation approach. Identifying approaches is therefore mainly an interpretive process relying on language used in the reports and description of techniques in both the reports and the interviews.

Table 5 highlights the best fit on an evaluation theory branch for each evaluation, based primarily on the methodology used and explicit statements about evaluation in the TORs and evaluation reports. The most populated branch is the use branch, followed by the methods branch and lastly the values branch. The reasons for placing each evaluation on a particular branch is described below, but in summary, the use branch is the most populated because of the focus on providing information to either the country office or the headquarters. The purpose of the

evaluations was to change the behaviour of the organization or partner organizations in some way either through program improvement or program replication. These evaluations primarily responded to the needs of intended primary users within the organizations. In the methods branch evaluations use was secondary feature that had at least some impact on the evaluation. There are widely different approaches in the evaluations within the methods branch but they do focus on demonstrating validity through the methodological approach. The values branch is the least populated. Although some of the evaluations on the other branches call on the evaluator to make value judgments on whether program elements were good or not, the use of valuing in the evaluations was generally limited. There was extremely limited focus on ensuring other stakeholders were involved in the evaluation process, reflecting the values of participation, democracy, and empowerment.

Some evaluation theorists Alkin and Christie have placed on the use branch would fit comfortably on the values branch. This is particularly evident with theorists who focus on the social justice, democratic involvement and empowerment dimensions of the evaluation. As such the work of Fetterman or the transformative participatory evaluation version of Cousin's and Whitmore's participatory approach would be at home on the values branch. What is striking in the evaluations in this study is those on the use branch do not draw on these approaches. So the values branch is not strongly represented in the sample and the notions of democratic involvement and empowerment through evaluations are missing.

Evaluation	Type of Crisis	Evaluation Theory Tree Branch(es)	Reasoning	Methodology
NRC-Uganda	Post Conflict Recovery	Methods	Uses processing tracing to test the theory of the intervention.	Process Tracing
DRC-Somalia	Ongoing conflict & Famine	Methods	Main objective asks evaluation to assess the performance of the project.	Cross Sectional Descriptive
Christian Aid-Pakistan	Natural Disaster (Floods)	Use	Appreciative learning approach focuses on learning lessons. Recommendations are a focal point of report.	Appreciative Inquiry
CARE -Gaza	Long term occupation/Conflict	Valuing	Uses MSC to place a value on the intervention. Stakeholders play a role in selection of MSC stories.	Most Significant Change
CARE-the Sahel	Recurrent Drought/Malnutrition	Use	The TOR clearly identifies the users of the report and these stakeholders were included in reflections workshop.	Qualitative methods
Support to Life-Turkey	Refugee Crisis-Fleeing Conflict	Use	Appreciative learning approach focuses on learning lessons. Considered RTE for improving ongoing work.	Appreciative Inquiry
Oxfam-South Sudan	Multiple-Low level conflict, post conflict, and food insecurity	Methods	The evaluator places a value on each of Oxfam's 10 global humanitarian indicators.	Desk Review
Oxfam-Kenya, Gaza & Haiti	Urban Food Security	Methods	Tests theory of change from urban interventions	Testing of theory of change
ACF-Philippines	Natural Disaster (Typhoon)	Use	Explicitly aimed at utilization by senior management at field and country office, and HQ level.	Real-time evaluation
Save the Children-Liberia	Refugee Crisis-Fleeing Conflict	Use	Aimed at use by STC at program improvement. The TOR does ask the evaluator to place value on the program, but use is more important.	Mixed

Table 5. Classification of evaluations on Alkin and Christie's evaluation theory tree.

Use

One of the recent major criticisms leveled at EHA is that evaluations simply were not used (Cosgrove & Buchanan-Smith, 2014) and the emphasis on use-oriented evaluations suggests NGOs have responded to that criticism. Use was usually directed towards organizational learning rather than empowerment of marginalized groups or cross-organizational learning. Responses from evaluators and NGO staff and the language used in the report suggest NGOs used the reports to provide some level of accountability to donors. In other words, NGOs were deliberately making themselves accountable to donors rather than donors holding NGOs to account, but some interviewees saw no evidence of donors actually using the evaluations. In this case, the reports provide evidence for accountability should the donors, as the main power holder, chose to use them.

Interviews with the evaluators and NGO staff gave a clear impression the primary intended users were country offices and headquarters.

“I think in this respect that is was for (the NGO) themselves. They were having a few issues at times with the intervention that they were making and I think it is always good for any organization to have a third party, and independent third party to come in and draw up some like win-win recommendations of how to improve things.”

“I think that the final evaluation was primarily an internal learning exercise not only for the country office, for the program staff, for the country office as well as for (the NGO) as an institution.”

Whether the main purpose of the evaluation was lesson learning or accountability, the evaluators and NGO staff generally identified the main users of the evaluation as being the NGO itself. Particularly with accountability focused evaluations, other users such as donors, the

general public and beneficiaries were mentioned. However, even with accountability focused evaluations, the accountability was often seen as being from the field to headquarters. It is difficult to identify the use approach with regard to the involvement of other stakeholders because they had little involvement in the evaluation and there was no concerted strategy within the evaluation to ensure these stakeholders used it.

Within the use branch, there are frequent elements of Patton's early approach focused on identifying intended primary users and working closely with them on the design, data collection and results interpretation. It is also possible to identify some of Cousins and Whitmore's ideas about organizational learning through the practical participatory evaluation (P-PE) approach in the evaluations. The involvement of staff in supporting the development of the evaluation, working on the collection of data, and analyzing and using the results shows elements of P-PE.

Many of the interviewees identified the close collaboration with headquarter, field staff, and buy-in from field staff, although the highest level of involvement came from INGO staff. Key internal stakeholders (staff) were identified prior to the evaluation and influential in designing the TOR and evaluation questions, consulting with the evaluation, supporting the data collection process and helping to interpret and negotiate the results. One of the recurring claims in the interviews with evaluators was ownership of the evaluations by the NGO staff would help support evaluation use in the future.

“the evaluation department had a kind of sense of ownership of it and therefore tried to take it to internal stakeholders as a product that they had confidence in”

“the debrief helped the ownership and helped identify or at least bring to everyone's attention issues around, which I thought was glaring”

The level and location of ownership varied. Some of the evaluators felt there was genuine buy-in to the need and purpose of the evaluation by the field staff and the country office and others felt headquarters were more invested in the results than the field staff. This has implications for the expected use of the evaluation. In general one can surmise ownership of the evaluation at the field level has the potential for more immediate program and project improvements, while evaluations which are primarily utilized by headquarters have more potential to impact global policy of the NGO.

One particular section of the use branch was missing from the sample. There was little evidence of use being conceptualized as being important for empowerment or social justice with beneficiaries being the main users of the evaluations. The idea of transformative participatory evaluation (T-PE) or of Patton's later ideas of utilization focused approach supporting the empowerment of recipients and the support of social justice (Patton, 1997) are not present, nor is there any indication Fetterman's empowerment approach (1994) is being used by the humanitarian sector.

It was notable from both the reports and the interviews that in the most immediate emergencies, which face bigger logistical constraints, there was less involvement of the field staff in the evaluation because of the time constraints placed upon them, particularly at the design stage. Save the Children's report explicitly identifies a limitation of the study because of the "stipulation in the consultant's terms of reference that the evaluation fieldwork should not interfere with emergency relief activities." Another evaluator concurred with this idea: "... [at] the field level they have so much other stuff to get on with that they probably don't have much time to spend on it. They understand—everyone understands—the need for it. No one says I'm not bothered with that I don't want to do that, that's a waste of my time. Everyone can see the

need for it but you know it's hard to get people to sit down and you end up talking all day and then someone finally gets back at 6 or 7 o'clock and you can see they have been working since 8 o'clock in the morning and then the evaluator says oh can we just talk for an hour." This evaluator identifies a major constraint to the participatory inclusion of staff members when the evaluation is being conducted during an emergency. The evaluator suggested field staff often accept the need for an evaluation but making time to be involved is a different matter. This provides evidence that based on field realities, evaluators will adapt to target users who are most in a position to utilize findings during the challenge of an immediate emergency.

Surprisingly given the focus in the evaluation purposes on use, only 2 of the 10 evaluations explicitly list the users of the report. It is however possible to interpret who intended users were in other evaluations from the language used, particularly in the TORs and stated purposes. The TOR for the ACF evaluation clearly identifies staff from ACF as target users, both at the country and HQ level. This is consistent with the stated users in ACF's evaluation policy (ACF International, 2011).

The intended users in three of the evaluations can be interpreted from statements in the report about the evaluation's purpose or intended use. The TOR for the CARE Sahel evaluation includes a paragraph on use of results in the desired methodology section. This mainly identifies CARE's global and national staff and offices as the intended users although there is a nod to accountability to the general public by stating the report will be publically available on CARE's website. Through the interpretation of the inception report (the only evaluation to publish the inception report) it is clear the evaluators also recognized CARE staff as the main users. The Christian Aid and Support to Life evaluations do not explicitly identify users but the TORs identify to whom the results are to be presented. The Support to Life TOR states: "For the

communication of results, an official report of the evaluation will be prepared and presented. This report focusing on practical recommendations will provide lessons to STL project staff and senior management for improvements in the continuation of STL's response related to the Syria crisis."

Methods

The evaluations placed on the methods branch are distinct from each other in approaches but are placed on this branch because they rely on methodological rigor to make the case for a valid evaluation. The more neo-positivist side of the methods branch was not present in the sample. None of the evaluations included in this study used experimental or quasi experimental designs. The use of experimental and quasi-experimental designs in EHA is extremely limited (Cosgrove & Buchanan-Smith, 2014). The difficulty and ethical constraints of identifying control groups combined with the limited amount of baseline data INGOs collect, and the fact most INGO staff and many evaluators are not specialized in the techniques used in experimental and quasi-experimental design explain this absence (Cosgrove & Buchanan-Smith, 2014). During general discussion about the practice of EHA rather than a sampled evaluation, one INGO staff member described a case where a donor had asked for control groups to be included in an evaluation but the INGO had argued it was impractical.

A small number of evaluations used statistical analysis as part of the evidence building process, but only one of the evaluations presents any finding in statistical language. The CARE evaluation in Gaza used regression analysis on pre-collected data from the project and a stress tolerance survey to identify areas of impact. This analysis was used to support the evidence of various themes of impact that developed from the stories of change. Limitations of the statistical data were described in the report: "the survey was an experimental instrument, without any

exploratory factor analysis done to see if these various statements fit together in terms of internal validity and reliability.” The report also noted a respondent bias problem that children might deliberately score poorly on a pre-intervention test because of a fear that if they scored well they would not be included in the program. In this evaluation, the statistical analysis is not the main driver of analysis, which comes primarily from the analysis of stories of change.

Two evaluations used detailed surveys as a major source of evidence. The NRC evaluation team contracted IPSOS Uganda, a polling firm, to support the quantitative data collection process and blended the findings into the qualitative analysis. Random sampling was used, but the lack of a control group and baseline data are raised as limitations of the evaluation. The DRC evaluation implemented two surveys. One used a convenience sample of the attendees at the food distribution point when the evaluation team was present. The other used random sampling but the evaluation team had to double the sample frame because of the difficulties of finding respondents. The data itself is presented in the format an opinion poll would be presented in a newspaper rather than how an academic paper might present its findings. There was no discussion of statistical significance, t-scores, factor analysis, and so on in either of the reports. The use of IPSOS by NRC, and some of the language used in the DRC report (for example, about data cleaning) suggests the evaluation teams in both of these evaluations had the necessary statistical knowledge to discuss their findings in more technical language, but chose not to do so. One explanation is this was partly recognition of the audience the reports were aimed at; that of NGO staff and possibly donors who may have limited statistical training.

Two evaluations on the methods branch placed a strong emphasis on testing intervention theories, and as such most closely align with a theory-driven evaluation approach (Chen, 2005). In both cases a key purpose of the evaluation was the generation of knowledge for use in other

programs around the world. It is perhaps surprising there was not more testing of theories of change and program theories in the evaluations, since evaluations based on theories of change are frequently discussed, particularly as some donors have moved from the traditional logic model approach to a theory of change approach (Cosgrove & Buchanan-Smith, 2014).

Two of the evaluations relied on an approach similar to the discrepancy model approach of Malcolm Probus. This approach relies on program managers or organizations deciding on objectives and standards for measurement, and the role of the evaluator is to identify discrepancies between the standards and the actual performance (Mathison, 2005). Oxfam's South Sudan program evaluation relies on this approach. The evaluation used a standardized protocol that Oxfam has set for its global humanitarian indicator tool (GHIT) evaluations. The GHIT has been developed to facilitate comparisons among countries and involves a desk based review of key project documents, secondary information and skype conversations with country office staff that allows the evaluator to give a score on a four point scale for 12 quality standards. This was the only evaluation that used a standardized protocol. It is the only evaluation in the sample where the requirement for a TOR was waived because the GHIT itself and information available about the GHIT provide a framework similar to a TOR. TORs for other evaluations using the GHIT are available and demonstrate a standardized approach for all Oxfam evaluations using this methodology. It is also the only evaluation that did not include field visits. Oxfam developed the benchmarks, the quality indicators, and the evidence that should be used to make the judgment. The role of the evaluator was to identify how the Oxfam program is doing based on the standards that have been set by Oxfam's headquarters using the evidence and indicators as a guide. Oxfam developed the GHIT,

“As part of a wider organisational undertaking to better capture and communicate the effectiveness of its work, Oxfam developed an evaluative method to assess the quality of targeted humanitarian responses. This method uses a global humanitarian indicator tool which is intended to enable Oxfam GB to estimate how many disaster-affected men and women globally have received humanitarian aid that meets establishes standards for excellence. Equally importantly, it enables Oxfam GB to identify the areas of comparative weakness on a global scale that require institutional attention and resources for improvement.” (Featherstone, 2012, p. II)

Oxfam is using a standardized tool to provide both rigor and comparability to their evaluation. The focus on producing a methodologically valid evaluation reflected in the statement is a key focus of theorists within the methods branch (Alkin, 2011). The last sentence above also shows attention to the need to learn and make changes as a result of the evaluation, which is one of the important steps in the discrepancy model (Mathison, 2005).

The DRC evaluation relies on elements of the discrepancy model. The main objective of the evaluation was “to conduct an end of project evaluation based on the overall performances against the principle objectives.” (HERALD Consultants, 2012, p. 5) The report focused on how the project performed against objectives set in the project proposal and includes a summary table of achievements against the principal objectives. The evaluation was hard to place on a particular branch because it appears want to fulfil many different functions.

“The evaluation was intended to provide information and an opportunity for learning and accountability purposes, for the donor but importantly also for the rights holders at community level. It was expected to generate relevant findings, lessons, and recommendations which will be shared with DRC, partners and local stakeholders. The

evaluation results were expected to be used to guide and inform future programming regarding a continued programming of wet food and cash transfer project in southern Somalia. The evaluation was also to assess the performance of the project against key parameters including the project's relevance, effectiveness, efficiency, sustainability, timelines of activity implementation, and its strengths and weaknesses. This information was to be shared with key DRC stakeholders; current donors as well as potential donors and non-governmental and governmental actors." (HERALD Consultants, 2012, p. 5)

Within this stated evaluation purpose there is a claim it will provide accountability to donors, partners and beneficiaries, and it will also generate lessons and recommendations for DRC, partners and local stakeholders, and the information will be shared with donors, potential donors, non-government and government actors. There is therefore a focus on ensuring the evaluation is utilized by a wide-variety of stakeholders. However, because of the stated principle objective, and the comparison of discrepancy between objectives set in the proposal (presumably by program managers and other DRC staff) and actual performance, the evaluation was judged to use the discrepancy model and sit on the methods branch.

There is evidence within the evaluations also of the real-world approach advocated by Rugh, Bamberger and Mabry (2012), an approach developed specifically for evaluators working in international development and humanitarian relief. It recognizes field realities evaluators often face: being called in late in a project, having to work on a limited budget, and where necessary data such as baseline surveys are not available. The approach also recognizes evaluators cope with political pressures from a number of different stakeholders. Although the final step of the approach focuses on helping clients to use the findings, the foundation of the approach is that a

sound methodological approach will ensure the evaluation's findings are valid even given the constraints imposed upon it. The approach therefore sits on the methods branch.

Although none of the evaluations apply this approach step by step, some evaluations rely in part on this approach, particularly in the early steps of the process that focus on the constraints. Save the Children's evaluation has a number of constraints including lack of baseline data, pressure from the INGO to "not interfere with emergency relief activities" (Karlsen, 2012, p. 5), a TOR requiring answers to 26 evaluation questions spread across 7 criteria and an additional requirement for recommendations on 11 topics by one evaluator in 21 days work. The evaluation does not follow the real world evaluation model exclusively. It relies more on a program theory approach but elements of the real-world model are seen in how the evaluator deals with the constraints. Scaling back the evaluation scope is illustrated in the report, from the 26 questions and 11 topics for recommendations to only 10 questions and the recommendations are based on topics addressed in the questions rather than additional topics included in the TOR. The consolidation of the questions comes from compressing some of the initial questions together and dropping ones that might be difficult to address in the limited time available. For example, the questions for the effectiveness criterion in the TOR included questions about how well the project met international response standards and how well it mainstreamed gender and environmental concerns into delivery, and these were not included in the evaluation. The TOR also had the following two questions: "To what extent were the intended outputs and results achieved in relation to targets set in the project proposal/logical framework" and "How effective and appropriate was the project approach?" In the evaluation itself, this was compressed to "How effective was the project approach to achieving the intended outputs and results?" (Karlsen, 2012, p. 13).

The Save the Children evaluation is an example of one that draws on elements from different branches. The methods branch is represented by the elements of the real world approach identified above. The final designation of the evaluation was on the use branch though because the recommendations are very clearly directed at utilization by the country and field offices for the purpose of program improvement, such as “separate admin/finance/HR functions at the sub-office level” and “build up New Yourpea sub-office” (Karlsen, 2012, pp. 24–25). This finding is supported by the large scale of Save the Children staff in the evaluation, which is demonstrated in the annexes by the lists of staff interviewed. The evaluation did not define a program theory model though and this contributed significantly to positioning it on the use rather than methods branch.

Valuing

The valuing branch is the least represented of the three branches, and only one of the evaluations sits on the valuing branch. This was CARE International’s evaluation in Gaza that used most significant change as its methodology.

Most significant change can be identified as an example of the participatory, possibly democratic approaches to evaluation. In some sense, the evaluation can be understood within the framework of MacDonald’s (MacDonald & Kushner, 2005) work on democratic evaluation where the evaluator supports the valuing of the project by a variety of different stakeholders and ensures the evaluation communicates differing opinions with a particular emphasis on hearing the voice of the most marginalized. In the case of this evaluation, the role of the affected community in valuing the project is limited by time constraints that meant only CARE staff and CBO leaders were involved in choosing the stories of change. Ideally this approach would have participants involved in that decision as well, but in this evaluation they were represented by the

CBOs rather than having direct voice. That said, this evaluation still sits within the valuing branch.

There are elements of the valuing branch in evaluations that sit on the other branches, but it is not the most significant element of the approach. Participation occurs mainly through the inclusion of beneficiary and community voices in the data collection stage. Value-orientated evaluators, particularly constructivists, concentrate on including diverse participant voices in an evaluation (Alkin, 2011). With the exception of Oxfam's evaluation of its South Sudan program, all the evaluations included beneficiaries only by collecting data through focus groups, community meetings and interviews. Many of the TORs placed strong emphasis on the need to ensure beneficiaries and other community members were consulted for their opinions in the evaluation, but this alone is not enough to qualify as a valuing approach. The involvement of community members is mainly an extractive process where they are only asked for data to answer specific questions. While the beneficiaries or communities may give responses to answer specific questions, the role of analyzing this data and determining the value of the program falls to either the evaluator or the INGO (i.e. the evaluator produces his/her findings and the INGO decides what these mean).

Additionally some of the evaluations place a stronger focus on whether the intervention met the needs of the affected communities than on whether the project achieved pre-determined goals laid out in a project proposal or log-frame. The Support to Life evaluation can be interpreted as defining the DAC criteria of relevance and appropriateness as what was relevant and appropriate to the needs of beneficiary populations and not the goals and objectives of either the donor or the initial program, and the evaluation questions for effectiveness ask how effect Support to Life was in delivering on those needs. This allowed the evaluation to address the

needs and position of the most vulnerable and powerless group, the Syrian refugees, rather than more powerful groups such as the NGO or donors. The evaluation therefore has strong links to House's idea of justice as a key criterion for evaluations (1980, 2014). Despite this, the evaluation still sits on the use branch because of its strong focus on use by Support to Life senior management and staff. In total, 5 of the evaluations identify relevance from the intended beneficiary perspective, 2 identify relevance as being focused both on beneficiary needs and donor or organizational goals, and 1 identifies relevance as whether the project met the goals of the funding drive. Two others, both of the Oxfam evaluations do not include relevance as a criterion. There is therefore evidence NGOs evaluators are identifying relevance through the lens of the affected community needs rather donor or NGO goals.

In general though there is little evidence valuing approaches are being utilized in current EHA practice. Although elements of ideas that values-based evaluators focus on are present in some evaluations, the strong focus is on methods or utilization. Only CARE International's evaluation in Gaza relies heavily on an approach that fits within the valuing branch. As with the use branch, with the exception of the evaluations that asked local civil society partners to be involved in analyzing findings, the ideas of democratic participation and community empowerment through evaluation are missing from this sample.

Who Benefits From Current Evaluation Practice?

The final sub-question of the research looked at who is benefitting from current evaluation and accountability practices, particularly focusing on whether affected communities benefit. One of the limitations of this research is that it did not empirically assess the impact and use of the evaluation. Most of the evaluators interviewed have only informal impressions of how

the evaluation had been used. Even the participants who were INGO staff had limited knowledge of the use and impact of the evaluation.

One question asked of the interview participants was, “Who do you think benefitted from the evaluation?” The participants recognized this was a subjective question and the answers were often speculative. “That’s the million dollar question,” one participant responded, identifying both the difficulty in answering that question and the importance of the question to INGOs, communities and the utility of the evaluation. The general belief of the interviewees was the INGOs themselves were benefitting the most from the evaluations. Within this perspective, there were differing opinions as to whether the benefits were mainly being seen by headquarters or country offices. Indirectly some participants felt affected communities were also benefitting. Additionally, some evaluators felt national level partners often benefitted, usually this was local civil society partners but national governments were also highlighted as potential beneficiaries of the evaluations.

INGOs

The perception INGOs themselves benefitted the most from the evaluations is understandable since the evaluations focused on lesson learning and internal accountability. The conclusion the majority of the evaluations sit on the use branch of the evaluation theory tree supports this idea. One respondent from an INGO’s headquarters said “ultimately the purpose of our evaluations for me is continuous learning and improvement of our programs and I think that is what most people in our programs want, we want it to be useful. To improve the way that we do our work and that is both at the project level and the organizational level”.

Although this response suggests NGO headquarters staff hope the evaluations are useful both at the field level and the headquarter level, there was disagreement amongst the evaluators

interviewed as to which level of the NGO the evaluations are useful. Some identified both headquarters and field offices whilst others specifically identified one or the other. One evaluator believed the operations staff in country benefitted from the evaluation whilst noting “I contacted (the headquarters) after the report was completed and submitted to ask so what did you guys do now and I never heard anything back.” On the other hand another evaluator stated “I would imagine the people in the countries to some extent shrug their shoulders... I wonder if it's much around organizational learning and potentially if there is some sort of donor spin on this ...potentially some of the donors might find it interesting, but it might be more a kind of headquarters temperature check.”

Donors

The direct influence of donors on the evaluations sampled was low. Although some evaluations were identified as fulfilling a requirement of accountability to donors, there was no evidence donors were actively involved in the design of evaluations. Interview participants suggested there was limited follow-up by donors from evaluation results. One NGO staff member suggested donor involvement was primarily in bigger strategic evaluations of entire programs.

However, the evaluations do provide evidence that might facilitate accountability for donors, should they choose to use the reports for this purpose. This provides greater accountability to donors than the affected communities who appear to be often denied access to the information at all. One should remember evaluations are only one of many documents that will be sent to donors. Regularly reports (usually quarterly), final reports and financial audits also complement the documentation that can be used for accountability.

Local Institutions

Local institutions were also identified by some respondents as having benefitted from the evaluations. Two participants identified the national government (or a particular department of government) as an entity benefitting from the evaluation: “the evaluation was supposed to also have a role in helping the Government structure their future programs.” And from the same evaluator: “I think we were able to contribute to the Government program as well, and I think that may have longer term benefits if that sticks around.” This occurred in projects where a handover of operations to the government was on-going or imminent and in projects/programs where the focus of work was transitioning from a humanitarian relief approach to a development approach.

Evaluations that involved local civil society groups provided the highest level of community representation. Some of the evaluators and NGO staff believed that often the main benefactor of the evaluation were the CBOs and local NGOs as the evaluation gave them the opportunity to reflect on their work and also to hold the INGO to account for their management of the project. One interviewee described how a meeting with local organizations allowed for the leaders of the organizations to question some of the actions of the INGO but also to identify the gains they had made for their community, consider important issues, and reflect on how to sustain the changes as the level of INGO support changed.

Affected Communities

Benefit to the affected communities was seen as much more indirect. A number of respondents identified the potential for affected communities to benefit from improvements to programs as a result of organizational learning by the NGO. One respondent identified the situation as “indirectly they would, because if the systems improve, if the operations improve,

therefore the beneficiary support improves. So it's more or less A+B=C situation." And another stated those who would benefit would be "people involved in the next disaster by a slow and hopefully uphill improvements in how we do these things." 'Hope' or 'may' were words used by 3 of participants when considering benefit to affected populations. This suggests a much less tangible benefit than those identified for the INGOs. All but one of the respondents who mentioned beneficiaries or the local community in their response saw the benefit as coming through the mechanism of program improvement. One respondent highlighted the ability to reflect on how the project has changed them as being of benefit to the project's beneficiaries. None of the respondents highlighted increased accountability, a knowledge gain that might help the community influence future projects, or the empowerment of the under-represented as being a mechanism by how the communities and beneficiaries benefitted from the evaluation.

There is some evidence in the evaluation reports to support the idea that affected communities benefit indirectly from the evaluations. There are references to the implementation of recommendations from previous evaluations, mid-term reviews and feedback from monitoring, suggesting INGOs used previous evaluative work to make programmatic changes. These changes were acknowledged and assessed in more recent evaluations. The Christian Aid evaluation included an assessment of how well Christian Aid and their partners have implemented recommendations from previous evaluations as an indicator within the efficiency, impact and effectiveness criteria of the evaluation. The assessment includes identifying how well suggestions that arose from community feedback had been incorporated into current practices; for example: "At partner level, good examples of lessons learned being incorporated into successive phases of house-building, for example Muslim Hands' iteration of house layout between different phases to reflect residents' preferences" (Morgan et al., 2013, p. 14). This

demonstrates the formative process of learning from the previous evaluation included understanding affected community concerns and responding to them. Another example of this is in the CARE Gaza evaluation which notes the changes made to programming as a result of a mid-term evaluation. “In response, however, to concerns noted in the E2FI midterm evaluation that graduates of the program may experience an “impact drop”, CARE WBG included in the E2FII design a six-month long “Graduate Club” to reinforce key skills and strengthen key outcomes from the program.” (Shah, 2013, p.20)

Not all evaluations considered previous evaluation recommendations and there are examples of those where management systems were not in place to effectively utilize evaluation learnings. CARE International’s evaluation in the Sahel was critical of the M&E system within the different country offices, arguing both that it reduced opportunities to improve service to beneficiaries through learning in Chad and criticizing the quality of a previous evaluation in Niger:

“CARE Chad’s M&E system was not well-developed. This inhibited CARE Chad’s ability to account more fully to donors, and its own learning on how to improve the quality and impact of its work to better assist beneficiaries” (Gubbels & Bousquet, 2013, p. 22)

“CARE Niger’s own evaluation report for the CT project does not mention the sample size nor the methodology used to reach its findings.” (Gubbels & Bousquet, 2013, p. 28)

This not only questions whether beneficiaries are benefitting indirectly, but also how broad the finding that INGOs are benefitting from evaluations is.

A majority of the evaluations included downward accountability as a criterion or indicator within a criterion as part of the evaluation. It could be argued beneficiaries benefit

indirectly from this as it allows organizations to hold field offices accountable for implementing these practices in the field. If, as argued in previous research, downward accountability mechanisms are beneficial to affected populations (Featherstone, 2013), then including downward accountability as a regular criterion in evaluations will help ensure these practices are carried out at a field level. As such an indirect benefit can be traced through this part of the evaluation, although no evaluations systematically evaluated the NGO's compliance with their accountability framework, as required by HAP benchmark 6.3.

Conclusion

The findings therefore suggest INGOs and local civil society partners of the INGOs are benefitting the most from current evaluation practice. What benefit there was for beneficiaries and affected communities is more indirect, coming through INGOs improving their operations as a result of evaluations and increasingly pressuring field staff to focus on downward accountability mechanisms by including downward accountability as a criterion in the evaluation. The benefit to donors is less clear but may come from the continuation of evaluation practices that have benefitted them in the past. The discussion section will address this further and analyze the implications of these findings.

Chapter 5: Discussion and Implications

There has been considerable discussion in the last decade and a half about downward accountability in humanitarian relief, the findings from this study suggest in practice such accountability is quite limited. The review of evaluation policies and accountability statements of the INGOs show at a policy level INGOs support the idea of using evaluation as mechanism to provide accountability to affected populations. However, I found for the most part affected community involvement in evaluations was only at the data collection stage. Communities are not involved in setting the scope of the evaluation and in most cases not included in analyzing and interpreting findings. Even access to information, a basic level of accountability, appears to be denied to communities, as the results of the analysis are not shared with them. Instead INGOs privilege internal lesson learning and benefits to the affected population from evaluation come indirectly from program improvement. The opportunity to influence decisions that affect the community is not provided by current EHA practice and so does little to alter the power imbalances within the humanitarian community.

How Are INGOs Operating in Humanitarian Relief Contexts Using Evaluations as a Mechanism to Provide and/or Assess Accountability to Affected Communities?

My research found there was limited use of evaluation to provide accountability to affected communities. Accountability provided to affected communities through INGO evaluation practices comes indirectly from program improvement and lesson learning. My research found no substantial evidence of accountability from information sharing, participation

and the opportunity to influence decisions. Accountability to intended beneficiaries during project implementation was assessed in 60% of the evaluations.

Accountability Through Lesson Learning and Program Improvement

A weak form of accountability is being provided by lesson learning and program improvement. For example, benchmark 6 of HAP's Standard in Accountability and Quality Management focuses on learning and continual improvement (HAP International, 2010a). There was evidence this benchmark was partially being achieved. The evaluators and NGO staff interviewed believed many of the evaluations contributed to the improvement of programs, a perception based on informal feedback rather than formally tracking the use of the evaluation. Nine out of ten evaluations included a learning component as part of the evaluation purpose, suggesting INGOs view evaluation as a mechanism for this kind of accountability. This interpretation is reinforced by the analysis of the implementation of previous evaluation findings suggesting many NGOs take responsibility for improving programs based on recommendations of evaluations. Lloyd, Blagescu, and Casas (2005) argue learning is the key connector between evaluation and accountability. Learning allows an organization to feed key information into its activities and ensure program improvement. They also argue broad based accountability should allow multiple stakeholders to influence decisions and this comes through participation in the learning process. Learning in an evaluation does not however offer accountability if stakeholders affected by the INGOs actions are not involved in the process. The INGOs in this study did not foster a participatory approach to learning and this weakens the accountability offered through the evaluations. Focusing on learning to provide accountability without ensuring input for multiple stakeholders covers only one element of accountability though, and because it does not

allow communities the opportunity to significantly influence decisions made about them, it is a very passive, weak version of accountability.

Accountability Through Information Sharing and Participation

Accountability to affected communities through the sharing of information was not present in most of the evaluations. Wenar (2006) has argued accountability is not just about being responsible but being seen to be responsible and this is missing from the evaluations sampled. Humanitarian aid providers retain control of the information shared with affected communities and despite attempts to improve communication through accountability initiatives, affected communities still feel uninformed about basic elements of projects (Anderson, Brown, & Jean, 2012). The organization may be accountable to its mission but if the community is not presented with the results of the evaluations, it is hard to argue organizations are being accountable to them.

There was no evidence of evaluations using a more active, participatory approach to accountability. Sharing information is an important element of accountability but without the opportunity to use this information to influence decision making the accountability is limited. The stakeholder approach to accountability relies on participation at all stages to allow affected communities to have meaningful input into decision making (Lloyd, 2005). This more active version of accountability would have the community either meaningfully using the information or participating in the making the evaluation judgment. The research found no participation of communities in setting the scope of the evaluation and very little in analyzing and negotiating the results. The idea communities could use information to influence decisions that affected them was not present in the selected sample. It could be argued consulting beneficiaries and other community members during the data collection stage allows individuals to influence future

decisions, but this is a very weak influence. The influence the affected community has is extremely limited because they are excluded from discussing what questions should be asked in the first place, and not participating in analyzing what the responses mean. This is consistent with evaluation approaches within the NPM framework that limits beneficiary and community involvement and thus opportunity to set evaluation standards, identify what is good and analyze what the findings mean (Norris & Kushner, 2007). Control is in the hands of senior managers and evaluation becomes a more technocratic endeavour than a democratic empowerment exercise (Chouinard, 2013). Power is not transferred and the opportunity to influence decisions is lost.

While accountability can occur without participation, participation enhances the meaningfulness of downward accountability. In his discussion of accountability mechanisms, Ebrahim (2003) argues participation allows the tool of evaluation to be used for downward accountability. Participation moves accountability from a passive version where affected populations are recipients of information about a program to an active version where communities' leverage is increased by systematic participation holding NGOs to account by deciding what to evaluate and what the results of an evaluation mean (Ebrahim, 2003). Participation allows accountability to become more than just access to information but provides the means to influence decision making (Blagescu et al., 2005). My findings suggest EHA evaluations are currently not being used as a mechanism for even the passive form of accountability, as there is little evidence of feedback to the communities.

The lack of participation means opportunities for affected communities to hold INGOs to account are limited. The distinction between holding oneself accountable and being held to account is important. Learning involves the INGO making itself accountable to communities

through program improvement. Participation provides both the opportunity for communities to hold INGOs to account by valuing and judging their actions and ensures learning incorporates their views in a meaningful manner (Blagescu et al., 2005). This gives communities the opportunities to be involved in something which has the power to make significant differences to their lives. Under the current humanitarian structure, affected communities have very little power to influence decision making as beneficiaries are often viewed by humanitarian agencies as passive recipients of aid (Cosgrove & Buchanan-Smith, 2014). Active participation in the project cycle at the design, implementation and assessment stages would help to redress some of these power imbalances by ensuring communities have a say in decisions that are made about them (Anderson et al., 2012; Featherstone, 2013).

Evaluation plays a critical role in the project cycle through judging the success of programs and improving program delivery, and it is particularly important for the community when interventions are on-going. The longer-term delivery of programs is more common because emergencies and disasters are becoming more protracted and the lines between development and humanitarian relief are increasingly blurred (Tammainga, 2011). So while a project may be ending, giving an evaluation a summative purpose, the delivery of programs may be on-going, giving the evaluation a formative purpose. With this scenario the active involvement of communities of holding NGOs to account rather than the passive receiving of information from evaluations the NGO chooses to give them, becomes more important for influencing decisions that affect the community. This emphasizes forward looking accountability rather than just a final project activity (Lloyd, 2005). Even if the project is coming to an end and the INGO is leaving the project location, the community's involvement in the evaluation has longer term benefits such as building evaluative capacity and having information to mobilize for

advocacy purposes with government officials and other NGOs. INGOs did not support the reduction of power imbalances by not using evaluation as a mechanism for accountability and the limited opportunities for influencing decisions are reduced further if communities are not involved in the design and use of evaluations.

Civil Society Involvement

The research found the evaluations with the highest level of accountability to affected communities were those involving local civil society partners in the evaluations. In these examples the involvement of these partners helped the feedback cycle of results to the affected community, and enhanced the democratic process by asking them to analyze what value to place on particular results, what important lessons should be learned, and how this matters moving forward. Ownership of the evaluation results is enhanced by being involved in at least some of the stages of the evaluation (Shula & Cousins, 1997; Green, 1988).

The selection of groups representative of the marginalized in society rather than entrenched local interests is an important consideration when working with local partners and empowerment or social justice oriented evaluation assumes responsibility for ensuring the most marginalized voices are heard (House, 1980; Segone & Bamberger, 2011). The selection of local partners was not analyzed in the evaluations and so it is difficult to definitively comment on this issue, but the findings suggest accountability to affected populations can be improved through the inclusion of local partners. Even these evaluations were not truly participatory as they did not include community members from all sections of society, but did provide more representation than the other evaluations in the sample.

Assessing Downward Accountability in Project Implementation

To some extent evaluations are being used as mechanisms to assess downward accountability to affected populations. While only Oxfam's evaluations significantly diverged from the DAC criteria, Oxfam and two other INGOs included downward accountability or its equivalent as a separate criterion. Research included in the 2009 HAP report found "only 13% of the evaluations were explicitly considering accountability to intended beneficiaries and local communities" (HAP International, 2009, p. 16). The 2010 report found this number had risen to 60%. However this number is skewed by the inclusion of 17 Inter-Agency Standing Committee (IASC) commissioned studies which 'almost all' considered accountability to intended beneficiaries (HAP International, 2010b, p. 57). Whilst it is positive that an influential agency is assessing accountability, the inclusion of so many evaluation reports commission by one agency in the sample affects the findings significantly. If ISAC evaluations were removed from the sample, the figure is closer to 40%. Sixty percent of my sample considered accountability to affected populations. With ACF included, 40% had a separate criterion for downward accountability. Another 20% built downward accountability into the DAC criteria; one in relevance and one in efficiency. This research and the findings in the HAP reports are not directly comparable but they do contribute to the conclusion there has been a gradual improvement in the use of evaluation as a mechanism to assess downward accountability.

However none of the evaluations used the standards of an accountability initiative to systematically assess downward accountability. None of the evaluations for example took HAP's benchmarks to analyze whether the project was compliant with them, and no other standards were comprehensively assessed either. Therefore evaluation is not being used as a mechanism to

systematically assess INGO performance against set standards of initiatives. Instead assessment is based on elements of accountability considered important by the INGO or project staff.

Policy vs. Practice

The lack of evidence of assessment against accountability initiatives and the failure of INGOs to adhere to their policies on participation at every stage of the evaluation cycle shows a disconnect between policy rhetoric and field practices. There has been an explosion of accountability initiatives in the sector (Alexander, 2013; Everett & Friesen, 2010; Mitchell & Knox-Clarke, 2011) and this can create confusion and disillusionment for field staff who are overwhelmed with requests to prove accountability (Mitchell & Knox-Clarke, 2011). In my research the evaluation closest to complying with policy requirements of the INGO was ACF's in the Philippines. ACF's evaluation policy makes no attempt to argue evaluations should be participatory or that potential users include affected communities. The policy requires a UFE approach focused on working with primary users (decision-makers) within ACF. The evaluation therefore met the requirements of the policy because the guidelines do not require beneficiary involvement in the evaluation. The other agencies' evaluation policies are much broader and inclusive. The evidence in my research shows policy requirements for community inclusion are not being followed.

The Listening Project identified increased levels of mistrust and feelings of being disrespected when power holders espouse an ideal at a policy level but are perceived to ignore these ideals on the ground (Anderson et al., 2012). By not following their own policies INGOs are signaling participation in the evaluation is not significant to them. To be useful to potential users, an evaluation has to be credible (Patton, 2009), and the credibility of an evaluation can be

undermined when the INGO's policies on participation are ignored. The potential for mistrust within the community seems high.

The Findings As a Whole

There are many other uses of an evaluation than just the provision of accountability to affected populations. The evaluations in this sample do have utility and seem to in general meet the purposes laid out in the TOR and draw logical conclusions from the evidence collected. The reports answer the evaluation questions and there was informal evidence from the interview participants that they were useful to the INGOs. Not all evaluations should be aimed at providing downward accountability directly and other approaches are at times more appropriate. There are many other purposes an INGO may prioritize in a particular evaluation such as testing intervention theories for transference to other locations. Nor is the use of evaluations for upward accountability wrong. Being accountable to government and individual donors large and small is important to the integrity of the aid sector. Additionally, the logistics of certain crises mean using the evaluation as a mechanism for downward accountability is not always relevant.

The sum of the whole findings from this study is more important than assessment of each individual evaluation. Even with the multiple uses of evaluation, one would expect some evaluations in this sample would focus on downward accountability, involve beneficiaries at more than just the data collection stage, and provide opportunities for communities to influence decisions being made about them. None of the evaluations included communities in designing them and only when there is CBO involvement was there any local representation in analyzing the results of the evaluation. This suggests the opportunity for evaluation to directly provide downward accountability is not being taken by INGOs. In each evaluation there may have been a good reason for the approach, but the sum of the whole findings show INGOs are missing or

avoiding an opportunity for communities to hold them to account. Within the aid sector, there are limited opportunities for affected communities to hold INGOs to account and to have an influence over decisions that are made about them. Evaluation can provide one of those opportunities and if the practice of EHA is not allowing this, then it is a missed opportunity from a narrow field of possibilities.

Where Does Power Lie?

The common accountability approach, where there was one, was the production of evidence those in power could use should they wish to, whether this was INGOs, either the country office or headquarters, or donors. One of the interview participants alluded to the idea that current practice of accountability in the humanitarian world focuses more on providing evidence power holders can use if they want to, than actually undertaking the active process of doing an evaluation with stakeholders, including the NGO and the project's beneficiaries. "It's a kind of gesture of accountability. Its providing a source of evidence that could be used should the power brokers choose to have a process of self-reflection and ultimately a sense of accountability for something they had a role in."

This raises the question of how power affects the practice of evaluation. House (1980) argued the role of the evaluator is to help reduce power imbalances by ensuring the voice of the marginalized is given most weight in an evaluation. He argued those who do not hold power are unlikely to have the views represented in deliberations of evaluation results and so an evaluator needs to address this imbalance by ensuring weight is given to the experiences of the powerless. The policy statements of four of the INGOs identify accountability as a means of reducing power imbalances, and the evaluation policies of the INGOs include statements identifying evaluation as providing downward accountability. Accountability to affected communities necessarily

involves the ceding of a certain amount of power because it allows the communities to influence decisions (Blagescu et al., 2005). The traditional understanding of power within the humanitarian system is donors hold the highest level of power followed by big INGOs and host governments, small INGOs and finally local civil society groups and affected communities holding the least amount of power (Collaborative Learning Projects, 2010; Darcy, 2013; HAP International, 2010b). So for evaluation to redress existing power imbalances, the voice of affected communities must be included.

I found little evidence to suggest current practice of EHA is altering the traditional balance of power. The main power imbalances my research was able to look at were between the affected communities and the INGOs. INGOs control the evaluation scope, questions, access to intended beneficiaries and community members, and dissemination of findings making the approach more bureaucratic (where the INGO is substituted for a government agency) than democratic (MacDonald & Kushner, 2005). In the bureaucratic approach, knowledge stays in the hands of the power-holders and is not disseminated to or shared with the marginalized. In most of the evaluations beneficiaries did not have the opportunity to control the direction of the evaluation or the use of the results. Even in the cases where local civil society groups were involved, the INGO still controlled the development of the evaluation and only through invitation were partners involved. It is hard to conclude the evaluations helped reduce power imbalances in the sector. This leads back to the conclusion that current practice of EHA evaluation does not give affected communities influence over decisions made about them nor recourse to action or sanctioning for bad performance.

Donors are perceived to hold the most amount of power in the humanitarian world (Collaborative Learning Projects, 2010; Darcy, 2013; HAP International, 2010b) and this

assumption explains many INGO practices. Actual donor involvement in the evaluations was low. Donor power comes from INGOs operating within a structure designed by donors and the desire for a product that can be useful to donors. The style of report production (Rugh & Bamberger, 2009), the use of DAC criteria (Chianca, 2008) and the evaluations in the sample that focus on analyzing performance against goals set out in a proposal (Feinstein & Beck, 2006) are all influenced by the desire of INGOs to maintain good relationships with the donor.

There is also a power differential between the INGO and the evaluator, and this is important for understanding how evaluations are practiced. In the sampled evaluations, the evaluator's range of approaches is constrained by the pre-determined limits of the evaluation. The option of a truly participatory approach seems typically forestalled because the INGO have already decided on the scope of the evaluation, the criteria for assessment and in many cases the questions that can be asked. The ownership of the evaluation results also lie with the INGO. Although INGOs do publish the results, they have control over how, where and when they are shared.

What Particular Evaluation Approaches Are Being Used With Which Forms of Accountability in Humanitarian Relief Work?

A goal of the research was to analyze what evaluation approaches are being used with particular forms of accountability. The answer to this question is limited by the findings showing very little accountability to affected communities and what accountability there is comes through learning and program improvement. As there is a fairly consistent approach to accountability, instead I analyzed what approaches are currently being used and identified what is missing from these approaches. The key missing element of the evaluations was the participation of affected

communities at stages other than during data collection. The lack of participation means approaches favouring community empowerment and democratic involvement were not used.

The analysis illustrates five of the evaluations should sit on the use branch, four on the methods branch and one on the values branch, based on Alkin & Christie's (2004) evaluation theory tree. Additionally, a number of evaluations on the methods branch had secondary elements of use. The lack of participation of affected communities meant the empowerment or democratic approaches within the use branch and approaches focused on valuing were missing from the sample. The theorists who could be associated with the evaluations on the use branch are those who identify use by the organization as primary. The theorists identified include Patton's early work on utilization and Cousins and Whitmore's P-PE approach. By excluding affected communities from meaningful participation, the evaluations on the use branch did not include work focused on empowerment and community use such as Cousins and Whitmore's T-PE approach or Fetterman's empowerment approach.

These findings can be understood when considering the conclusion that the type of accountability to affected communities offered by current EHA practices is a limited, indirect version of accountability via lesson learning. Had a broader version of accountability that included affected populations having access to information and actively participating in judging the project had been found in the research, then one would expect to see more evaluation approaches focused on empowerment and democratic valuing. If this were the scenario more evaluations would have been on the values branch, and the approaches employed in the evaluations on the use branch would have drawn more on ideas of empowerment.

The power imbalances between the INGO and the evaluator help explain these missing elements. The control of the INGOs in setting the scope and often questions for the evaluation,

how the results are disseminated, and the limited timeframes for the evaluations push the evaluations into the methods and use branch and away from the valuing branch, and empowerment sub-branch of use. Evaluation practice in this scenario becomes focused on traditional hierarchical notions of accountability (Edwards & Hulme, 1996) and the technocratic ideals of NPM (Chouinard, 2013). This has the effect of narrowing the field of potential users. To maximize use, a potential user needs to feel invested in the evaluation (Patton, 2009), and by excluding communities from meaningful participation in the evaluation the field of potential primary users narrows to INGO staff and donors. In the absence of a structure allowing or encouraging evaluators to be more inclusive and participatory, the evaluations focus most on methodological rigour and ensuring use by the INGO.

The missing elements in the evaluations are instructive in understanding the control and direction of current EHA practice. The INGO focus on internal lesson learning privileges working primarily with staff as stakeholders and producing methodologically rigorous evaluations that can be shared globally. The missing elements of full and democratic participation contribute to reducing the use of evaluation to provide accountability to affected populations.

Structural Problems

Why has this happened? NGOs have policies that favour downward accountability and it is reasonable to assume both NGO staff and evaluators do care about providing the best possible service to affected communities. Policy talk has also focused on enhancing humanitarian accountability, and yet INGOs seem to be missing the opportunity to use evaluation to enhance this goal. Structural problems probably account for a lot of these difficulties.

Time. One of these structural problems is time, a recurring problem in EHA (Polastro, 2011). The time available to do the evaluation is limited and so in most cases involves a very restricted amount of field observation and data collection. Not all of the evaluations specified schedules but from those that did, we see an average time available for the evaluations was 15-20 days and of that time usually a third, but not more than half, was spent in the field including travel days. This limits exposure to communities and beneficiaries. In many cases the evaluation team consists of more than one individual thus expanding the total days available, but with the exception of NRC's evaluation in Northern Uganda none of the evaluations had a large team: one had a team of 3 people, four had a team of 2, and four had only one evaluator. INGO staff were often involved in supporting data collection as well but whilst possibly increasing the efficiency of data collection, this does not increase the evaluators' exposure to communities.

One of the reasons for time constraints in evaluation is budgetary, and limited budgets reduce the impact of evaluations (Twersky & Arbreton, 2014). EHA is not unique and suffers from the same budget constraints as evaluation in other sectors (Bamberger, Rugh, & Mabry, 2012). However, time constraints do not just occur because of budget limitations. Planning also impacts the time allocations in an evaluation. Participation requires a different approach to planning and utilization of time. The traditional model of desk based planning followed by a short period of field data collection followed by writing and briefing headquarters on the results is not appropriate for a participatory evaluation approach. Polastro (2011), citing examples from Mozambique and Pakistan, argues good planning can significantly contribute to increasing time spent with beneficiaries in RTEs. This approach requires a reallocation of time from other tasks. It requires INGOs to facilitate evaluators spending more of the contracted time in the field, which would mean more of the planning stage is spent with beneficiaries. An alternative to the

evaluator taking this role would be ensuring project staff are invested in developing questions with the community. Particularly in a fast-moving, complex emergency ensuring there is a staff member with the necessary experience and training to undertake this task can be difficult. Research of field staff has shown resistance to new initiatives because they feel overwhelmed by multiple requests from senior management along with the pressures of implementing program activities (Anderson et al., 2012; Beattie, 2011)

My findings show evaluators are precluded from employing truly participatory approaches by the time allocated to them in the evaluation and the expected schedules for using that time. Even the evaluation that offered the most participatory approach, CARE's evaluation in Gaza, still had a very limited timeframe which reduced the community participation in the selection panels to the local CBOs. Most significant change (MSC) ideally would include a selection panel of the beneficiaries themselves but time did not permit this, nor did the timeframe allow for participation in the development of the evaluation questions. Unless INGOs rethink how they plan evaluations and refine their expectations for the use of external evaluator's time, the current time budgets allocated to evaluations will continue to be inadequate for ensuring participatory approaches.

Where evaluations sit within the project cycle. The small amount of time available for the evaluations suggests INGOs are not prioritizing evaluations within their projects. The evaluations appear to be seen as distinct from the rest of the project cycle. This reduces the opportunity for evaluations to influence decisions about the operations within a country (Hallam, 2011) which weakens the potential for accountability to the affected communities. Most of the evaluations were conducted in situations where there was potential for formative learning, whether it was for an on-going project or learning for future programming in the country. Even

the evaluations that required a summative judgment on the project asked for formative learning from the evaluation as well. Only one evaluation was conducted by an evaluator who had been involved earlier in the project in an evaluative capacity; in this case the same evaluator had also conducted the mid-term evaluation. Another was conducted by an evaluator who had evaluated the country program a number of years earlier. The evaluators are not involved in supporting the design of the project or the development of the M&E system but are coming in for a short period of time to conduct an evaluation that is independent of any previous evaluator involvement.

Building partnerships between evaluators and organizations can support participatory evaluations by helping stakeholder commitment to evaluations (Mathison, 1994). Genuine participation from community members would also help build a trusting relationship between the community and the evaluator, therefore increasing acceptance and ultimately the accountability of the evaluation.

This approach allows organizations to benefit from the relationship without the cost of a permanent, developmental focused external evaluation. It prevents the loss of independence donors value so highly that may occur if the evaluation becomes purely internal (Mathison, 1994). Building relationships does require a genuine commitment to reflective learning and genuine participation of affected communities to provide the accountability envisioned by Lloyd, Blagescu and Casas (2005). Without this the INGO can be guilty of building the cozy insider relationships identified by Perouse de Montclos (2012).

A number of evaluators had evaluated programs of the INGO in other countries but only one had been involved in the particular project at an earlier date. Instead many of the evaluators face the real world evaluation issues identified by Bamberger, Rugh and Mabry (2012). They are called in late in a project to conduct an evaluation under time pressure where they had not set up the M&E system. The pressure to both fit as much as possible into a proposal and for project

staff to implement as many activities as possible contributes to the feeling of evaluation as a stand-alone activity, rather than one built into the entire project-cycle. This reduces the opportunities to build relationships with community members and so diminishes the possibility to the community contributing to the formative learning elements of the evaluation. This mitigates the opportunity for the community to influence future programming.

Understanding participation. Another barrier evaluators face is the understanding of participation and how it facilitates accountability. Aid recipients do not see their input in ‘participatory approaches’ making any difference in decisions that affected them (Anderson et al., 2012). Four of the TORs refer to participation at the same time as presenting pre-determined evaluation questions defined without any community input, giving a limited time-frame to complete the evaluation and thus interact with the community, and requiring a product that is not accessible to the community. Save the Children’s TOR asks for “Consultations/interviews with a sample of beneficiaries (including children) to ensure meaningful participation of beneficiaries in the evaluation.” (Save the Children, 2012, p. 4). This evaluation schedule gave enough time for six days of field visits for data collection. It is hard to argue this actually allows for ‘meaningful participation’, a point the evaluator agreed with as she was critical of the “lack of investment in systems and tools for promoting a higher degree of stakeholder engagement in program design, monitoring and evaluation” (Karlsen, 2012, p. 9) and further arguing the M&E data is ‘extractive’ and “the active participation of beneficiaries and stakeholders in the project management cycle is still on an “as needed” basis.” (Karlsen, 2012, pp. 9&10).

Evaluators employing a narrow reading of participation would consider collecting data from beneficiaries to be a participatory approach (Cullen & Coryn, 2011). Other evaluators would focus on the participation of management and decision makers. Evaluators using a P-PE

approach would fit into this category (Cousins & Whitmore, 1998). The more empowerment focused participation described in the Save the Children manual (O'Neill, 2012) though is of the type envisaged in T-PE (Cousins & Whitmore, 1998), democratic evaluation (MacDonald & Kushner, 2005) or deliberative democratic evaluation (House, 2005). This evaluation frames participation as inclusive and continuous for all groups throughout the evaluation, but Karlsen (2012) identifies this type of participation as being missing from Save the Children's M&E system in Liberia. The same was true of the approaches used by most evaluators in this sample.

Even the inclusion of beneficiaries in data collection has not been guaranteed in EHA practice in the past (Cosgrove & Buchanan-Smith, 2014; HAP International, 2008). In 2003, 52% of evaluations assessed in ALNAP's Review of Humanitarian Action were judged to be poor in the consultation and participation of beneficiaries. In 2010, HAP's annual report found 78% of the sampled evaluations interviewed beneficiaries (HAP International, 2010b). This research found 90% of the evaluations consulted beneficiaries to some degree. The findings are not directly comparable as ALNAP's review looked at participation as well as consultation. Participation in this research was found to be far less present than consultation. Similarly the sample for the HAP annual report includes UN agencies, IASC, and donor evaluations as well as INGOs. However this fits in the narrative of beneficiary consultation in evaluations slowly improving. This may be a necessary first step on the road to improving the use of evaluation as a mechanism for downward accountability. The improvement in interviewing beneficiaries does suggest progress, although in a limited way. It is still a very token form of participation (Ebrahim, 2003) and not one which provides meaningful opportunity for influencing decisions about the program.

Evaluation products. A further constraint the TORs place on the evaluators is the type of product they are asked to produce. Basic advice to evaluators in reporting is to consider the needs of the stakeholders (Alkin, 2011; Torres, 2005). Evaluation TORs that request just a report in English, French or Spanish suggest the INGO is only considering the needs of senior managers, headquarters and donors (Rugh & Bamberger, 2009). All the TORs ask for a final written report and six of TORs detailed the headings the report should be structured around. By contrast, none of the TORs ask for the production of material that could be used to present the findings to the local communities and none ask the evaluator to involve the community in developing innovative a culturally appropriate presentation, which would be a sign the results are available in an appropriate format (Segone & Bamberger, 2011). As noted by one evaluator who in a previous evaluation designed comic strips to present the findings to the community, strategies for presenting the results in a more accessible format than a dense report in English are available but by and large they are not used. TORs require written reports from the evaluators and the time available for producing them is limited, which no doubt seriously limits opportunities to produce other reporting formats a community might find accessible.

Here again the historical power structures of the humanitarian world play a part. INGOs want a product that can be shared for accountability purposes, even if the evaluation's main focus is not accountability, and the common belief is donors want a structured report in English, Spanish or French. A report in English, Spanish or French is seen as a good product to have, as it increases opportunities for learning within an organization and cross-organizational learning. However it comes at the expense of the production of material suitable and useful for affected communities, and even the translation of a report is rare and the budget available is often cited as a reason for this (Rugh & Bamberger, 2009). None of the interview participants were aware of

the evaluation reports being translated. Translating the report into the local language would increase its utility for local civil society organizations and other literate local stakeholders who might be able to access it.

The interview participants could not be fully sure how the results had been used after the contract had finished because control was in the hands of the INGO. However, the common belief was the results had not been shared in any meaningful manner. Just asking evaluators to produce more material appropriate to the local context, would probably mean the utilization of the evaluation by the community would be higher even if the evaluator did not lead the feedback process as it would provide over worked field staff with a readily available tool or strategy to use. As it is, in most evaluations, evaluators are asked to provide a traditional report conceptualized when evaluations were very clearly aimed at accountability to donors.

Moving beyond the immediate effects this has on the affected community, the requirement of writing an evaluation report in a Westernized reporting format in English, French or Spanish along with the lack of any empowerment element within the practice of EHA, does not contribute to the development of an indigenous evaluation community in the affected countries. In my sample there is a noticeable Westernized cadre of evaluators. Most are either originally from or currently based in Western countries. The aid world in general privileges French and Spanish, and particularly English (Grey, 2013; Taylor-Leech, 2013). Senior staff in country offices are usually required to speak and write a Western language. The requirement to produce evaluation reports in formal styles may be a barrier to entry for local evaluators. This reduces the long-term possibilities to build evaluative capacity in local communities and ultimately to de-colonize aid. Globally there are increasing numbers of evaluators working in their indigenous contexts who may be well positioned to provide a bridge for meaningful,

participatory evaluations. For example, the African Evaluation Association (AfrEA), now in existence for about a decade, promotes “made in Africa evaluation.” Another example is the International Organization for Cooperation in Evaluation’s (IOCE) “Peer-to-Peer program”, launched in 2013, aims to strengthen local capacity. 32 national and 6 regional voluntary professional organizations participated in this scheme (IOCE, 2014).

Methodologies. Despite these structural constraints, there was evidence evaluators can influence the evaluation in a way that makes a difference. The CARE’s evaluation in Gaza shows evaluator’s can persuade INGOs of the effectiveness in using more innovative methodologies to provide more participatory evaluations that offer accountability to affected populations. The evaluation demonstrates it is possible to produce work donors value as well as providing information to the community. Use of MSC stories is not suggested in the TOR and indeed the objective of the evaluation was “to assess whether the set targets and anticipated results of the project implementation during the two year of the project life were achieved” (Shah, 2013, p. xxxiv (annexes)), but the methodology used focused on assessing unintended as well as intended results. MSC has been used as a methodology which strengthens both learning and downward accountability through participation (Davies & Dart, 2005). It includes the voice of the affected communities to understand what actually happened in a program rather than focusing on pre-determined objectives, and improves the feedback mechanism through consultation on the selected stories of change (Shah, 2014).

In CARE International’s evaluation, the stories of change were translated into Arabic and a brochure of the stories was a product of the evaluation for future use by the CBOs. The English version of the brochure is included in the report at appendix A (Shah, 2013). This complimented the more traditional evaluation report to be shared with donors. In this example, the product still

requires literacy to be able to utilize the brochure but since the literacy rate of the population over 15 in Gaza and the West Bank in 2012 was 96% the production of written material in the local language is much more appropriate here than it would be in CARE's Sahel evaluation where the literacy rate in Chad was 37% (The World Bank, 2014). The stories of change written in Arabic may not be used by the local community, but at least the opportunity for them to do so exists because they have access to the information.

This was the only evaluation to use a methodology involving the community in both collecting and analyzing data. The context of the evaluation matters. It was conducted in a location where access to the beneficiaries was possible, where time is not taken by travel and at the time security constraints did not affect the evaluation. It is a project those who identify humanitarian relief in the more traditional sense of the immediate saving lives and relief of suffering would consider to be a development project. These characteristics may have given the evaluator more flexibility than evaluations in the other contexts. The MSC technique requires organizational buy-in because it involves field staff coordination with CBOs and beneficiaries. When TORs stress the importance of not interfering with field operations the use of such innovative approaches may be discouraged. The INGO purpose for doing the evaluation, the control over the design and the limited exposure of the evaluators and NGO staff to these methods probably explain why only one evaluation used a methodology that including the community in both data collection and analyzing the results.

Evaluators did affect the evaluations through prioritizing of particular questions or what was important in the relevance criterion. The evaluator in Save the Children's evaluation had to reduce the number of questions included in the TOR because of time considerations and chose to include questions focused more on beneficiary need and opinion in the TOR rather than

questions that looked at global standards or indicators. The evaluators in DRC's evaluation did not focus on the relevance of the project to the policies of DRC and its donor as stated in the TOR but on the opinions of beneficiaries on the appropriateness of the service provided.

While these interventions may make a small contribution to improving the consideration of affected community concerns, they do not alter the big picture of limited participation and the lack of accountability to affected populations. The choice of methodology can also only do so much. It cannot alter the lack of community consultation in the design of the evaluation if the questions have already been decided and when the timeframe is so constrained.

The DAC criteria. The continued dominance of the DAC criteria also creates structural barriers to using the evaluation to provide accountability to affected communities. Since the initial design of the development criteria in the early 1990s and the adaptation for complex emergencies in the late 1990s there have not been official updates of the criteria. Beck, Ramalingham and Cosgrove (2009) argue the DAC criteria reflect the weaknesses in the humanitarian sector identified by the international community at the time they were developed. The criteria may therefore exclude topics that have gained more recent importance within the sector. For example, if the DAC criteria were re-written today they would include accountability as a criterion (Cosgrove et al., 2009). Resilience is another concept that has become increasingly important as it provides a link between humanitarian aid and development work (Kindra, 2013; Tamminga, 2011) and potentially would replace connectedness and sustainability. The lack of change may reflect the mammoth bureaucratic effort to create them in the first place and a belief they remain a huge step forward from previous approaches (Chianca, 2008). The criteria still have a strong influence on evaluation policies and were used as a structure for many of the evaluation reports in this sample.

In ALNAP's RTE guide, Beck, Ramalingham and Cosgrove (2009) offer alternatives to the DAC criteria, including using a quality/accountability initiative, evaluating against an ideal type, and evaluating a specific issue. My research found most INGOs are not doing this, preferring to use the DAC criteria, with perhaps the addition of one or two organization specific criteria. The use of the DAC criteria for organizing an evaluation is not necessarily problematic but the over-reliance on them may stifle other innovative approaches (Kaplan, 2014). One effect of relying on the DAC criteria, can be to dilute findings on downward accountability because of confusion over which criterion this concept fits into (HAP International, 2009). Some INGOs are partially changing this by including downward accountability as a criterion and this probably has some impact on INGOs take greater responsibility for accountability within projects. However it is not universally applied by INGOs. Perhaps given the dominance of the DAC criteria within the sector, updating them to include a separate criterion for downward accountability would have considerable impact on both evaluation practice and project implementation. Oxfam was the only organization that really moved away from the DAC criteria and has included accountability in its evaluations.

The dominance of the DAC criteria within the humanitarian community suggest that traditional sources of power remain with donors more than other stakeholders. One of the arguments for using the DAC criteria is they provide uniformity among evaluations and allow comparisons between interventions (ACF International, 2011; Cosgrove et al., 2009). This idea privileges learning at a global level rather than the community level. The use of the DAC criteria alone does not preclude participatory methods or approaches that focus on accountability through empowerment and democratic involvement. The automatic use of them though can stifle

possibilities for using different approaches or encouraging innovation to tackle new concerns that have arisen in the sector since the DAC criteria were first devised.

Who Is Benefitting?

The final sub-question of the research asked who is benefitting from current evaluation practices in EHA. INGOs were the main beneficiaries of current practice identified by the interview participants but local civil society partners also benefitted. Donors were considered to benefit because of the production of evidence they could use for accountability purposes, should they chose to do so. The benefit to affected communities was much less direct. The benefits identified came through the improvement of programs rather than a direct benefit of being able to influence decisions that are made about them.

INGOs

The interview participations had differing views on where the main ownership of the evaluation lay within the INGO. This has implications for the use of the evaluation and ultimately who benefits as stakeholders who feels ownership in the evaluation will be more likely to use it (Patton, 2009). There will be a more direct impact on project implementation when ownership is felt at the country or project level than if the ownership is more keenly felt by the headquarters. Ownership of the evaluation by headquarters will be more likely to have a policy level impact and increase the potential for diffusion of intervention strategies and approaches to other country programs. When analyzing the location of ownership against the classification of evaluations on the evaluation theory tree we can conclude the methods evaluation approaches are the evaluations where ownership appears to rest more at headquarters. Most of the evaluations classified within the use and value branches appeared to have a higher level of ownership at the country office level. This is not surprising. With the exception of

Oxfam's South Sudan evaluation, the evaluations were conducted where the intervention takes place. This ensures staff members who participated in the evaluation were much more likely to be from the country office. It is therefore easier for the evaluator to focus on utilization if those expected to use the evaluation were present to participate. This is not the case for those evaluations owned more by headquarters, leading to approaches more suitable to the methods branch. Who benefits within the INGO depends on who feels ownership of the evaluation and the findings from my research find the benefits to be split between headquarters and country offices.

Affected Communities

The benefit affected communities gain from current evaluation practice will be interpreted differently depending on one's perspective on organizational learning, the purpose of evaluation and the humanitarian sector in general. If one assumes organizational learning comes from participatory reflection of key internal staff members, then current practice may be delivering benefits to affected communities. By engaging in reflection and developing actionable outputs, evaluations offer the opportunity for INGOs to learn from experiences and improve their program delivery which ultimately benefits affected populations. This conclusion would require greater research to trace the effect of evaluations on the performance of INGOs at a local and global level. Many INGOs identify learning from past performance as being a key element of being accountability. Taking responsibility for past actions and ensuring learning feeds into future decisions and helps ensure an organization is accountable. This idea also guides one of HAP's six benchmarks on accountability.

I would argue evaluation can offer more than this. It can be an opportunity for community empowerment, social justice, and the re-aligning of power imbalances by giving

marginalized populations a genuine voice in decision making and program improvement (MacDonald & Kushner, 2005; McGee & Gaventa, 2011). The benefit to affected populations from current practice is weak because INGOs are missing or avoiding the opportunities evaluation offers. Current practice includes the voice of the affected community at only one stage of the evaluation and therefore democratic involvement is by and large missing. Limited participation in the evaluation process reduces ability of affected communities to hold INGOs to account (Blagescu et al., 2005). Throughout a project cycle there are only limited opportunities for affected communities to actively influence decisions that impact them. Participation does not necessarily provide accountability and accountability can be provided without participation. However, given the contexts the projects were delivered in and the large power imbalances in the system, participation offers the most direct means of providing accountability (Blagescu et al., 2005). Participatory approaches to project design, to ongoing monitoring and through complaints mechanisms offer windows of opportunity for communities to influence decisions. Evaluation is another window of opportunity and current practice appears to be closing off this opportunity for this most part. Taking away this opportunity reduces the number of potential moments for influence and this impacts the level of accountability given to affected communities.

The evaluation profession and the humanitarian sector have the tools available to use evaluations to provide accountability to affected communities (Cullen & Coryn, 2011). Considerable work has been done in both the humanitarian relief and international development sectors on developing participatory tools designed to improve the influence of local communities on decisions that affect them. For example, CARE International was able to use MSC to increase the involvement of CBOs in the democratic process of assessing a project's performance. This evaluation provided opportunities for both upward and downward accountability, even if in this

case the benefits and accountability are to the CBO more than the affected population in general. The evaluation providing opportunities for the CBOs to hold CARE to account, and to learn about the impact of their work and the opinions of their communities but still provided CARE with a product that could be used for donor accountability. The tools therefore exist to allow communities to use evaluations as an opportunity to hold INGOs to account for their actions, and strengthen their influence in the decision making process. My findings suggest these tools are currently not being utilized to any large extent.

Concluding Thoughts

The challenge for the evaluation community moving forward is how to ensure evaluations provide utility to INGOs in program improvement while ensuring affected communities have the opportunity to influence decisions that affect them. The structure of the humanitarian system is built around significant power imbalances and evaluation on its own cannot redress these imbalances. A considerable re-think of the system is needed to level out these differences. This would require affected communities to be genuinely engaged in decision making processes and given the power to control projects and sanction performance. Within the current structure, effective evaluation practice does give options for redressing the power imbalances by a small amount. Even without the more radical idea of placing control for commissioning and running evaluations in the hands of the affected community, employing evaluation approaches giving the community more say in the process would greatly improve the utility of evaluation as a downward accountability mechanism. To achieve this, INGOs need to be more inclusive at the start and end of the evaluation. Communities need to be given genuine opportunities to set the scope of the evaluation, analyzing data and have access to the results. INGOs also need to create more opportunity for evaluators to engage communities and ensure

evaluators are enabled to use innovative methodologies to achieve this. Doing this could ensure learning opportunities from evaluations continue and INGOs are genuinely being held to account by the people they work for.

Limitations and Future Research

Limitations

The clearest limitation to this research is that it focuses on only one accountability mechanism. In addition to evaluations, mechanisms such as complaints systems, ongoing feedback, reports, professional associations, and participatory steering committees are mechanisms for different versions of accountability. The purpose of this research was to identify how evaluation is being used as an accountability mechanism but does not address how INGOs are using other accountability mechanisms.

A further limitation is the research could only sample evaluation reports publically published. One of the main criticisms leveled against INGOs is the reluctance to publish evaluation reports. The efforts of organizations such as ALNAP and HAP International has improved this situation and certain organizations such as CARE International commit to publish all their evaluations, but it is estimated only a fraction of evaluation reports are published and the INGOs often only publish positive evaluation findings (Pérouse de Montclos, 2012). The evaluations sampled may potentially have only given a partial picture of the situation in reality. The interviews with field practitioners helped mitigate this issue to a certain extent but the potential the sample gave a less than full picture remains.

Tracing the full design and use of the evaluations was outside of the scope of the research. Understanding how the evaluation was designed, questions set, and the results negotiated and disseminated to stakeholders came from an interpretation of the reports and

discussion with NGO staff and evaluators. Many of the reports give limited information on these issues. Particularly in the case of the discussions with external evaluators, how the findings were used and whether they were presented to beneficiary communities was often based on informal follow up and impressions the evaluators gathered during their consultancy. It is therefore potentially possible the results are being shared more with the local communities than the research found. However this limitation is mitigated to an extent by including some NGO staff as interview participants.

I advocate strongly in the discussion section for participation of affected communities. Participation does have limitations though. There are contexts in which participation is not possible, and some where it can have negative effects (Hilhorst, 2002). The research was unable to look fully at the implications of increased participation in the particular context of each evaluation. The criticism of the limited attention to participation by INGOs is focused on the results as a whole rather than any individual evaluation in particular.

I did not attempt to filter out evaluations from the sample based on whether I judged the project or program to fit more within the development field than the humanitarian relief sector. Organizations and practitioners who espouse a narrow definition of humanitarian relief would probably take issue with the inclusion of some of the projects in this sample. I take the standpoint that the separation between humanitarian relief and international development is narrowing. Crises are becoming more common place, international dynamics are politicizing relief efforts, and with the increased impact of climate change, future resilience work is going to diminish the differences between the two sectors. I'm not making a judgment on whether this is a positive or negative development. My standpoint is simply this is happening and so the humanitarian sector needs to adapt to this. For this reason if the INGO who posted the evaluation on ALNAP's site

for humanitarian evaluations considered the work to have humanitarian elements to it, I considered this to be a reasonable justification for inclusion in my sample.

Future Research Directions

More research is needed in three main areas; the demonstration of benefit to communities from providing accountability through evaluation, more in-depth research into the process for designing, analyzing and disseminating evaluations in a participatory way, and comparing evaluation practices of organizations who publish reports with those who do not.

Limited research has been done to demonstrate accountability to affected populations is beneficial and leads to program improvement despite this being generally accepted within the sector (Featherstone, 2012). One of the assumptions of my research is the use of evaluation as an accountability mechanism to affected communities is beneficial to the community. More research demonstrating this link would support meaningful accountability through evaluation. Field staff are considered to be ‘action-orientated’ and demonstrating a process works is likely to be more effective in obtaining their buy-in rather than just proscribing a hypothetical concept that has not been proved.

Research looking in a more in-depth way at the process of using participatory evaluation approaches in humanitarian settings is needed. One potential research route is case studies demonstrating how and if a participatory approach supports accountability to affected populations, and increases mutual learning and the contextual features for successful implementation in contexts where humanitarian relief takes place. This would provide INGOs with research demonstrating how good practices in this field support their stated policy goals.

Comparative research of the INGOs who publish reports practice evaluation compared to the practices of those who do not is needed. World Vision does not publish its evaluation reports

but has a policy of taking evaluation findings back to the affected community with an emphasis on the community being able to use these findings (Huddle, 2012). Research is need to identify whether or not publishing reports frees up evaluators to use more innovative reporting approaches relevant to affected communities, rather than the structured formal donor-style report found in this sample. Comparing an organization that does not publish its evaluation results to one that does would help in understanding whether the drive for greater accountability through report publishing has actually hampered accountability to affected populations.

References

- ACF International. (2011). *Evaluation policy & guideline*. Retrieved from http://www.actionagainsthunger.org/sites/default/files/publications/Evaluation_Policy_Guideline_Enhancing_Organisational_Practice_through_an_Integrated_Evaluations_Learning_and_Accountability_Framework_06.2011.pdf
- Agyemang, G., Awumbila, M., Unerman, J., & O'Dwyer, B. (2009). *NGO accountability and aid delivery*. Retrieved from <http://www.accaglobal.com/content/dam/acca/global/PDF-technical/sustainability-reporting/rr-110-001.pdf>
- Alexander, J. (2013). Accountability 10 years in review. Has the balance of power shifted? *2013 Humanitarian Accountability Report*, 18–59. Retrieved from Humanitarian Accountability Partnership website: <http://www.hapinternational.org/resources/hap-publications.aspx>
- Alkin, M. C. (2004). *Evaluation roots*. Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412984157>
- Alkin, M. C. (2011). *Evaluation essentials: From A to Z*. New York, NY: Guilford Press.
- Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree. In M. Akin (Ed.), *Evaluation roots* (pp. 13–66). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412984157>

ALNAP. (2003). *Evaluating humanitarian action*. Retrieved from
http://ec.europa.eu/echo/files/evaluation/watsan2005/annex_files/ALNAP/ALNAP1 - Evaluating Humanitarian Action.pdf

ALNAP. (2006). *Evaluating humanitarian action using the OECD-DAC criteria An ALNAP guide for humanitarian agencies*. Retrieved from <http://www.alnap.org/resource/5253.aspx>

ALNAP. (2012). *The state of the humanitarian system*. Retrieved from
www.alnap.org/pool/files/alnap-sohs-2012-lo-res.pdf

Anderson, M. B., Brown, D., & Jean, I. (2012). *Time to listen. Hearing people on the receiving end of international aid*. Retrieved from
<http://www.cdacollaborative.org/media/60478/Time-to-Listen-Book.pdf>

Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real world evaluation. Working under budget, time, data and political constraints* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Barnett, M. (2011). *Empire of humanity: A history of humanitarianism*. Ithaca, NY: Cornell University Press.

Barry-Shaw, N., & Jay, D. O. (2012). *Paved with good intentions: Canada's development NGOs from idealism to imperialism*. Black Point, Nova Scotia: Fernwood Publishing.

Beattie, K. (2011). NGO accountability: Findings from Sudan. *Humanitarian Exchange*, October(52), 44–46. Retrieved from <http://www.odihpn.org/humanitarian-exchange-magazine/issue-52/ngo-accountability-findings-from-south-sudan>

Binder, A., & Meier, C. (2011). Opportunity knocks: why non-Western countries enter humanitarism and how to make the best of it. *International Review of the Red Cross*, 93(884), 1135–1149. doi:10.1017/S1816383112000409

Blagescu, M., Casas, L. D. Las, & Lloyd, R. (2005). *Pathways to accountability*.
http://oneworldtrust.org/publications/doc_view/210-pathways-to-accountability-the-gap-framework?tmpl=component&format=raw

Borton, J. (2004). The joint evaluation of emergency assistance to Rwanda. *Humanitarian Exchange*, (26), 14–18. Retrieved from <http://www.odihpn.org/humanitarian-exchange-magazine/issue-26>

Burns, J. (2010). Humanitarian aid v. national interests: The current dilemma in Pakistan. *The Georgetown Public Policy Review*. Retrieved from
<http://gppreview.com/2010/10/20/humanitarian-aid-v-national-interests-the-current-dilemma-in-pakistan/>

Cameron, M. A., Tomlin, B. W., & Lawson, B. (Eds.) (1998). *To walk without fear : the global movement to ban landmines*. Don Mills, On: Oxford University Press, Canada. Retrieved from <http://www.ebrary.com>

CARE International. (2008). *CARE international evaluation policy*. Retrieved from
<http://www.care-international.org/accountability>

CARE International. (2010). *CARE International humanitarian accountability framework policy statement and guidance note*. Retrieved from

<http://water.care2share.wikispaces.net/file/view/CARE+Humanitarian+Accountability+Framework+Pilot+Version+Feb+2010.pdf>

Charity Navigator. (n.d.). *How do we rate charities accountability and transparency?* Retrieved April 11, 2014 from

<http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=1093#.VFk46PnF-So>

Chen, H.-T. (2005). Theory-driven evaluation. In S. Mathison (Ed.), *The Encyclopedia of evaluation* (pp. 416–420). Thousand Oaks, CA: Sage Publications, Inc.

doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558.n542>

Chianca, T. (2008). Development evaluations : An assessment and ideas. *Journal of MultiDisciplinary Evaluation*, 5(9), 41–51. Retrieved from

http://www.ipdet.org/files/Publication-The_OECD-DAC_Criteria_for_International.pdf

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*, 34(2), 237–253. doi:10.1177/1098214013478142

Christian Aid. (2012). *Christian Aid's accountability framework*. Retrieved from
<http://www.christianaid.org.uk/Images/Christian-Aid-Accountability-Framework-Dec-2012.pdf>

Christie, C. A., & Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation*, 34(3), 131–135. doi:10.1016/j.stueduc.2008.07.001

Collaborative Learning Projects. (2010). *The listening project issue paper: Structural relationships in the aid system, (March)*. Retrieved from
<http://www.cdacollaborative.org/media/48009/The-Listening-ProjectIssue-Paper-Structural-Relationships-in-the-Aid-System.pdf>

Cornwall, A., Lucas, H., & Pasteur, K. (2000). Introduction: accountability through participation: developing workable partnership models in the health sector. *IDS Bulletin*, 31(1), 1–13.

Cosgrove, J. (2007). *Synthesis report: Expanded summary. Joint evaluation of the international response to the Indian Ocean tsunami*. Retrieved from
www.alnap.org/pool/files/Syn_Report_Sum.pdf

Cosgrove, J., Beck, T., & Ramalingam, B. (2009). *Real-time evaluations of humanitarian action-An ALNAP guide: Pilot version*. Retrieved from <http://www.alnap.org/resource/5595>

Cosgrove, J., & Buchanan-Smith, M. (2014). *Evaluation of humanitarian action pilot guide*. London: ALNAP. Retrieved from <http://www.alnap.org/eha>

Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close*. Charlotte, N.C.: Information Age Pub.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 1998(80), 5–23. doi:10.1002/ev.1114

Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7–22.
doi:10.1177/1098214012464037

Crotty, M. (1998). *The Foundations of social science research* (2nd ed.). Thousand Oaks, CA: Sage.

Cullen, A., & Coryn, C. L. S. (2011). Forms and functions of participatory evaluation in international development: A review of the empirical and theoretical literature. *Journal of MultiDisciplinary Evaluation*, 7(16), 32–47. Retrieved from
http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/288/312

DAC/OECD. (1999). *Guidance for evaluating humanitarian assistance in complex emergencies* Retrieved from <http://www.oecd.org/development/evaluation/2667294.pdf..>

Darcy, J. (2013). Have we lost the plot? Revisiting the accountability debate. *2013 Humanitarian Accountability Report*, 4–17. Retrieved from Humanitarian Accountability Partnership website: <http://www.hapinternational.org/resources/hap-publications.aspx>

Davies, R., & Dart, J. (2005). *The 'most significant change' technique: A guide to its use*. Retrieved from <http://www.mande.co.uk/docs/MSCGuide.htm>

DfID. (2011). *Saving lives, preventing suffering and building resilience: The UK Government's humanitarian policy*. Retrieved from
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67468/The_2

OUK_20Government_s_20Humanitarian_20Policy_20-_20September_202011_20-_20Final.pdf

DfID. (2013). *International development evaluation policy May 2013*. Retrieved from <http://www.oecd.org/derec/unitedkingdom/DFID-Evaluation-Policy-2013.pdf>.

Dionne, K. Y. (2014, March 28). Will an organization receiving US funds get away with discriminatory hiring practices? *The Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/monkey-cage/wp/2014/03/28/will-an-organization-receiving-u-s-government-funds-get-away-with-discriminatory-hiring-practices/>

Ebrahim, A. (2003). Accountability in practice: Mechanisms for NGOs. *World Development*, 31(5), 813–829. doi:10.1016/S0305-750X(03)00014-7

Edwards, M., & Hulme, D. (1996). Too close for comfort? The impact of official aid on nongovernmental organizations. *World Development*, 24(6), 961–973.

Everett, J., & Friesen, C. (2010). Humanitarian accountability and performance in the Théâtre de l’Absurde. *Critical Perspectives on Accounting*, 21(6), 468–485. doi:10.1016/j.cpa.2010.01.014

Featherstone, A. (2012). *Evaluation of Oxfam’s South Sudan Humanitarian Response*. Retrieved from <http://www.alnap.org/resource/11873>

Featherstone, A. (2013). *Improving impact: Do accountability mechanisms deliver results?* Retrieved from Humanitarian Accountability Partnership website:

<http://www.hapinternational.org/pool/files/improving-impact-do-accountability-mechanisms-deliver-results.pdf>

Feinstein, O., & Beck, T. (2006). *Handbook of evaluation*. (I. Shaw, J. C. Greene, & M. M. Mark, Eds.) (pp. 536–558). London: Sage.

Fetterman, D. (1994). Empowerment evaluation. *American Journal of Evaluation*, 15(1), 1–15.
doi:10.1177/109821409401500101

Fetterman, D. (2005). Empowerment evaluation. In Mathison S. (Ed.), *Encyclopedia of evaluation* (pp. 125–130). Thousand Oaks, CA: Sage Publications.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558.n165>

Fetterman, D. (2010). Empowerment evaluation. In M. Segone (Ed.), *From policies to results. Developing capacities for country monitoring and evaluation systems* (pp. 278–288). New York: UNICEF. Retrieved from <http://mymande.org/index.php?q=content/policies-results>

Fetterman, D., Rodriguez-Campos, L., Wandersman, A., & O’Sullivan, R. G. (2013). Collaborative, participatory, and empowerment evaluation: Building a strong conceptual foundation for stakeholder involvement approaches to evaluation (a response to Cousins, Whitmore, and Shulha, 2013). *American Journal of Evaluation*, 35(1), 144–148.
doi:10.1177/1098214013509875

Fournier, D. M. (2005). Logic of evaluation-working logic. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 239–243). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558>

Gibbs, D. N. (2009). *First do no harm*. Nashville, TN: Vanderbilt University Press.

Gibson, C. C., Andersson, K., Ostrom, E., & Shivakumar, S. (2006). *The Samaritan's dilemma: The political economy of development aid*. Oxford: Oxford University Press.

Global Humanitarian Assistance. (n.d.-a). Defining humanitarian assistance | Global Humanitarian Assistance. Retrieved February 18, 2014 from
<http://www.globalhumanitarianassistance.org/data-guides/defining-humanitarian-aid>

Global Humanitarian Assistance. (n.d.-b). Workstream: Delivery. *Development Initiatives*. Retrieved April 29, 2014 from
<http://www.globalhumanitarianassistance.org/workstream/themes/delivery>

Goetz, A. M., & Jenkins, R. (2005). *Reinventing accountability: making democracy work for human development*. New York, NY: Palgrave MacMillian.

Gray, R. H., Owen, D., & Adams, C. A. (1996). *Accounting and accountability, changes and challenges in corporate social and environmental reporting*. Upper Saddle River, NJ: Prentice Hall International.

Greene, J. G. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, 12(2), 91-1 16

Grey, A. (2013, October 17). We do aid, not English. [Blog post]. *Language on the move*. Retrieved November 11, 2014 from <http://www.languageonthemove.com/language-globalization/we-do-aid-not-english>

Gubbels, P., & Bousquet, C. (2013). *Independent evaluation of CARE's response to the 2011–2012 Sahel humanitarian crisis*. Retrieved from <http://www.alnap.org/resource/9703>

Hacsi, T. A. (2000). Using program theory to replicate successful programs. *New Directions for Evaluation*, 2000(87), 71–78. doi:10.1002/ev.1183

Hallam, A. (2011). *Harnessing the power of evaluation in humanitarian action : An initiative to improve understanding and use of evaluation*. Retrieved from
<http://www.alnap.org/resource/6123>

HAP International. (2005). *The humanitarian accountability report 2005*. Retrieved from
<http://www.hapinternational.org/resources/hap-publications.aspx>

HAP International. (2006). *The humanitarian accountability report 2006*. Retrieved from
<http://www.hapinternational.org/resources/hap-publications.aspx>

HAP International. (2007). *The 2007 humanitarian accountability report*. Retrieved from
<http://www.hapinternational.org/resources/hap-publications.aspx>

HAP International. (2008). *The 2008 humanitarian accountability report*. Retrieved from
<http://www.hapinternational.org/resources/hap-publications.aspx>

HAP International. (2009). *The 2009 humanitarian accountability report*. Retrieved from
<http://www.hapinternational.org/resources/hap-publications.aspx>

- HAP International. (2010a). *Guide to the 2010 HAP standard in accountability and quality management*. Retrieved from <http://www.hapinternational.org/pool/files/guide-to-the-2010-hap-standard-printer-friendly-version.pdf>
- HAP International. (2010b). *The 2010 HAP standard in accountability and quality management*. Retrieved from <http://www.hapinternational.org/what-we-do/hap-standard.aspx>
- HAP International. (2010c). *The 2010 humanitarian accountability report*. Retrieved from <http://www.hapinternational.org/resources/hap-publications.aspx>
- HAP International. (2011). *The 2011 humanitarian accountability report*. Retrieved from <http://www.hapinternational.org/resources/hap-publications.aspx>
- Haque, M. S. (2007). Revisiting the new public management. *Public Administration Review*, 67(1), 179–182. Retrieved from <http://www.jstor.org.ezproxy.library.ubc.ca/stable/4624553?seq=2>
- HERALD Consultants. (2012). *Final evaluation report of the wet feeding and cash transfer project in Southern Somalia*. Retrieved from <http://www.alnap.org/resource/7771>
- Hilhorst, D. J. (2002). Being good at going good? Quality and accountability of humanitarian NGOs. *Disasters*, 26(3), 193–212. doi:10.1111/1467-7717.00200
- Hilhorst, D. J. (2005). Dead letter or living document? Ten years the Code of Conduct for disaster relief. *Disasters*, 29(4), p351-369. doi: 10.1111/j.0361.2005.00297.

- Hofmann, C., Roberts, L., Shoham, J., & Harvey, P. (2004). *Measuring the impact of humanitarian aid. A review of current practice* (Vol. 44). Retrieved from <http://dspace.cigilibrary.org/jspui/handle/123456789/22658>
- Hood, C. (2001). Public management, new. In Smelser, N. J. & Baltes, P. B. (eds). *International Encyclopedia of the Social & Behavioral Sciences* (pp. 12553–12556). Amsterdam: Elsevier. doi:10.1016/B0-08-043076-7/01180-3
- House, E. (1980). *Evaluating with Validity* (2nd ed.). Beverly Hills, CA: Sage Publications, Inc.
- House, E. (2005). Deliberative democratic evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 105–109). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558.n138>
- House, E. (2014). Origins of the ideas in evaluating with validity. *New Directions for Evaluation, Summer*(142), 9–16. doi:10.1002/ev
- Huddle, J. (2012). World Vision: It has opened our eyes. In *Ignite your engagement: collaborative, participatory, and empowerment issues explored through ignite presentations*. World Vision. Retrieved from <http://igniteshow.com/videos/world-vision-it-has-opened-our-eyes>
- Human Rights Watch. (2009, September 22). *Sri Lanka: World leaders should demand end detention camps*. Retrieved April 23, 2014, from <http://www.hrw.org/news/2009/09/22/sri-lanka-world-leaders-should-demand-end-detention-camps>

Human Rights Watch. (2010). *"I want to help my people"* State control and civil society in Burma after Cyclone Nargis. Retrieved from

<http://www.hrw.org/sites/default/files/reports/burma0410webwcover.pdf>

International Red Cross and Red Crescent. (2014). *Code of conduct for the International Red Cross and Red Crescent and NGOs in disaster relief*. Retrieved March 09, 2014, from [http://www.ifrc.org/Global/Publications/disasters/code-of-conduct/Code of Conduct Version 2.pdf](http://www.ifrc.org/Global/Publications/disasters/code-of-conduct/Code%20of%20Conduct%20-%20Version%202.pdf)

IOCE. (2014). *History*. Retrieved November 18, 2014 from <http://www.ioce.net/en/history.php>

Jean, I., & Bonino, F. (2013). *We are committed to listen to you. World Vision's experience with humanitarian feedback mechanisms in Darfur. ALNAP/CDA case study*. Retrieved from <http://www.alnap.org/resource/8851>.

Jones, S., & Picanyol, C. (2011). *Mutual Accountability: Progress since Accra and issues for Busan*. Retrieved from <http://www.opml.co.uk/sites/opml/files/Mutual%20Accountability%20Study.pdf>

Jordan, L. (2009). *Mechanisms for NGO accountability*. Retrieved from Global Public Policy Institute website:

http://www.gppi.net/fileadmin/user_upload/media/pub/2005/Jordan_2005_Accountability.pdf

Kaplan, J. (2014). *DAC criteria*. Retrieved November 10, 2014 from
http://betterevaluation.org/evaluation-options/dac_criteria

Karlsen, M. (2012). *Save the Children emergency response to the Ivorian refugee crisis in Liberia: Final evaluation report*. Retrieved from <http://www.alnap.org/resource/6369>

Kennedy, D. (2004). *Reassessing International Humanitarianism*. Princeton, NJ: Princeton University Press. doi:978-0-691-12394-3

Kindra, J. (2013, March 4). Understanding resilience. *Irin*. Retrieved from <http://www.irinnews.org/report/97584/understanding-resilience>

Kinsella, E. A. (2006). Hermeneutics and critical hermeneutics: Exploring possibilities within the art of interpretation. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 7(3). Retrieved from <http://nbn-resolving.de/urn:nbn:de:0114-fqs0603190>

Kugu, B., & Oksak, E. (2013). *Evaluation study Support to Life's response to Syrian refugee crisis, 2012 & 2013*. Retrieved from <http://www.alnap.org/resource/10304>

Kushner, S., & Rotondo, E. (2012). Evaluation voices from Latin America: Editors' notes. *New Directions for Evaluation*, 2012(134), 1–4. Retrieved from <http://www.eblib.com>

Larner, W. (2000). Neo-liberalism: policy, ideology, governmentality. *Studies in Political Economy*, (63), 5-25. Retrieved from <http://spe.library.utoronto.ca/index.php/spe/article/view/6724/3723>

Lenn, M. (2006). *Dochas NGO Accountability : Issues , Lessons and challenges for*. Retrieved from http://www.dochas.ie/Shared/Files/4/ngo_accountability_paper.pdf

Lloyd, R. (2005). *The role of NGO self-regulation in increasing NGO accountability*. Retrieved from http://coddeconduitaong.ro/wordpress/wp-content/uploads/2011/08/The_Role_of_NGO_Self-Regulation_in_Increasing_Stakeholder_Accountability.pdf

Macauslan, I., & Phelps, L. (2012). *Oxfam GB emergency food security and livelihoods urban programme evaluation final report*. Retrieved from <http://www.alnap.org/resource/7074>

MacDonald, B., & Kushner, S. (2005). Democratic evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 110–114). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558.n141>

Mathison, S. (1994). Rethinking the evaluator role: Partnerships between organizations and evaluators. *Evaluation and Program Planning*, 17(3), 299–304. Retrieved from http://ac.els-cdn.com.ezproxy.library.ubc.ca/0149718994900094/1-s2.0-0149718994900094-main.pdf?_tid=4723f23c-6c28-11e4-aaf9-00000aacb35d&acdnat=1415988277_b49b6575603dbd86de66e1e84469cb50

Mathison, S. (2005). Discrepancy evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 118–119). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558.n150>

Mathison, S. (2008). The difference between evaluation and research. In *Fundamental issues in evaluation* (pp. 183–196). New York: Guilford Press.

Mathison, S. (2009). Mary e. corcoran lecture minnesota evaluation studies institute, 2009, 1–23.

Mathison, S., & Ross, E. W. (2002). The hegemony of accountability in schools and universities. *Workplace: A Journal of Academic Labor*, 5(1). Retrieved from
<http://louisville.edu/journal/workplace/issue5p1/mathison.html>

McGee, R., & Gaventa, J. (2011). *Shifting power? Assessing the impact of transparency and accountability initiatives* (pp. 1–39). Brighton. Retrieved from
<http://www.ids.ac.uk/files/dmfile/Wp383.pdf>

Melamed, C. (2014). Development myths-ODI's take on Gates letter. *Development Progress*. Retrieved June 03, 2014, from
<http://www.developmentprogress.org/blog/2014/01/22/busting-myths-development>

Mitchell, J., & Knox-Clarke, P. (2011). Reflections on the accountability revolution. *Humanitarian Exchange*, October(52), 3–5. Retrieved from
<http://www.odihpn.org/humanitarian-exchange-magazine/issue-52/reflections-on-the-accountability-revolution>

Morgan, N., Naz, S., & Sanderson, D. (2013). *Pakistan floods response 2010-12 evaluation for Christian Aid and its partners*. Retrieved from <http://www.alnap.org/resource/8819>

Munyas Ghadially, B. (2013). Programme accountability guidance pack: A Save the Children resource. Retrieved from <http://resourcecentre.savethechildren.se/library/programme-accountability-guidance-pack-save-children-resource>

Myers, M. D. (1995). Dialectic hermeneutics: a theoretical framework for the implementation of information systems. *Information Systems Journal*, 5(1), 51–70. doi:DOI: 10.1111/j.1365-2575.1995.tb00089.x

Najam, A. (1996). NGO accountability: a conceptual framework. *Development Policy Review*, 14(4), 339–353. doi:10.1111/j.1467-7679.1996.tb00112.x

Norris, N., & Kushner, S. (2007). The new public management and evaluation. In N. Norris & S. Kushner (Eds.), *Dilemmas of engagement: Evaluation and the new public management. advances in program evaluation* (Vol. 10, pp. 1–16). Oxford, UK: Elsevier. doi:: [http://dx.doi.org/10.1016/S1474-7863\(07\)10001-6](http://dx.doi.org/10.1016/S1474-7863(07)10001-6)

Norwegian Refugee Council. (2005). *Evaluation policy*. Retrieved from <http://www.eldis.org/go/home&id=23338&type=Document#.VBnWrPldWS0>

Norwegian Refugee Council. (2008). *Evaluation handbook: Learning from experience*. Oslo: Norwegian Refugee Council.

Norwegian Refugee Council. (2014). *Open information and aid transparency*. Retrieved November 04, 2014, from <http://www.nrc.no/?aid=9153669>

Nutt, S. (2011). *Damned nations: greed, guns, armies & aid*. Toronto: McClelland & Steward Ltd.

O'Neill, K. (2012). *Evaluation handbook*. Retrieved from Save the Children website: <http://resourcecentre.savethechildren.se/library/evaluation-handbook>

Obrecht, A., Laybourn, C., Hammer, M., & Ray, S. (2012). *The 2011 / 12 DEC accountability framework assessment evaluation and peer review process*. Retrieved from
<http://www.oneworldtrust.org/component/content/article/93-showcase/360-the-201112-dec-accountability-framework-assessment>

OECD. (2010). *Evaluation development cooperation*. Retrieved from
www.oecd.org/dac/evaluationnetwork

OED. (2013). Accountability, n [Def.1]. In Oxford English Dictionary Online. Retrieved February 12, 2014, from
<http://www.oed.com.ezproxy.library.ubc.ca/view/Entry/1197?redirectedFrom=accountability#eid>

Okumu, W. (2003). Humanitarian international NGOs and African conflicts. *International Peacekeeping*, 10(1), 120–137. doi:10.1080/714002390

Overseas Development Institute. (2007). *Voice for accountability: Citizens, the state and realistic governance*. Retrieved from <http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/576.pdf>

Oxfam. (2010). *Oxfam policy on program evaluation*. Retrieved from
http://www.oxfam.org/sites/www.oxfam.org/files/file_attachments/story/oxfam-program-evaluation-policy-dec10.pdf

Patton, M. Q. (1997). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage Publications Inc.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Thousand Oaks, CA: Sage Publications Inc.

Patton, M. Q. (2009). Utilization-focused evaluation. In M. Segone (Ed.), *Country-led monitoring and evaluation systems Better evidence, better policies, better development results* (pp. 252–277). Geneva: UNICEF. Retrieved from <http://www.ceecis.org/remf/Country-ledMEdsystems.pdf>

Pérouse de Montclos, M.-A. (2012). Humanitarian action in developing countries: who evaluates who? *Evaluation and Program Planning*, 35(1), 154–60.
doi:10.1016/j.evalprogplan.2010.11.005

Polastro, R. (2011). Real time evaluations: contributing to system-wide learning and accountability. *Humanitarian Exchange*, October 20(52), 10–12. Retrieved from <http://www.odihpn.org/humanitarian-exchange-magazine/issue-52>

Riddell, R. (2007). *Does foreign aid really work?* Oxford: Oxford University Press.

Roberge, J. (2011). What is critical hermeneutics? *Thesis Eleven*, 106(1), 5–22.
doi:10.1177/0725513611411682

Rocha Menocal, A., & Sharma, B. (2009a). *Citizens' voice and accountability: Understanding what works and what doesn't work in donor approaches. Lessons and recommendations emerging from a joint donor evaluation*. Retrieved from ODI website: <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3767.pdf>

- Rocha Menocal, A., & Sharma, B. (2009b). *Evaluation of citizens' voice and accountability*. Retrieved from ODI website: <http://www.odi.org.uk/projects/120-evaluation-citizens-voice-accountability>
- Rodríguez-Campos, L. (2012). Advances in collaborative evaluation. *Evaluation and Program Planning*, 35(4), 523–8. doi:10.1016/j.evalprogplan.2011.12.006
- Rodriguez-Campos, L., & Rincones-Gomez, R. (2012). *Collaborative evaluations: Step-by-step*. Stanford, CA: Stanford Business Books.
- Rogers, P. J., Petrosino, A., Huebner, Tracey, A., & Hacsi, T. A. (2000). Program theory evaluation: Practice, promise and problems. *New Directions for Evaluation*, 2000(83), 5–13. doi:10.1002/ev.1177
- Rugh, J., & Bamberger, M. (2009). Real world evaluations. Conducting quality evaluation under budget, time, data and political results. In M. Segone (Ed.), *Country-led monitoring and evaluation systems Better evidence, better policies, better development results* (pp. 200–237). Geneva: UNICEF. Retrieved from <http://www.ceecis.org/remf/Country-ledMEdsystems.pdf>
- Salgado, R. S. (2008). European money at work: Contracting a European identity? *Sciences Po, April*. Retrieved from <http://ssrn.com/abstract=1522240>
- Santamaria, C. (2013, September 18). After CIDA, AusAID: Australia “integrates” aid into foreign affairs. *The Development Newswire*. Retrieved from

<https://www.devex.com/news/after-cida-ausaid-australia-integrates-aid-into-foreign-affairs-81826>

Save the Children. (2012). SCUK Liberia final evaluation report annexes. Retrieved from
<http://www.alnap.org/resource/6369>

Segone, M., & Bamberger, M. (2011). *How to design and manage equity-focused evaluations*. New York, NY: UNICEF. Retrieved from
http://mymande.org/sites/default/files/EWP5_Equity_focused_evaluations.pdf

Scriven, M. (2005). Logic of Evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 118–119). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.eprozy.library.ubc.ca/10.4135/9781412950558.n322>

Scriven, M. (2007). *Key evaluation checklist*. Retrieved from
http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf

Shah, R. (2013). *Eye to the future II: Final external evaluation report*. Retrieved from
<http://www.alnap.org/resource/10244>

Shah, R. (in press). Assessing the 'true impact' of development assistance in the Gaza Strip and Tokelau: 'Most Significant Change' as an evaluation technique. *Asian Pacific Viewpoints*

Shula, L. M. & Cousins, J. B. (1997). Evaluation use: Theory, research and practice since 1986. *American Journal of Evaluation* 18 (3), 195-208. DOI: 10.1177/109821409701800302

Tamminga, P. (2011). Sustainability in humanitarian action. In “*Sustainable humanitarian action: bridging relief to development*” (pp. 1–6). Dara International. Retrieved from <http://daraint.org/wp-content/uploads/2011/12/Sustainability-in-humanitarian-action.pdf>

Taylor-Leech, K. (2013, October 13). English and development work [Blog post]. Retrieved November 11, 2014 from <http://www.languageonthemove.com/language-globalization/english-and-development-aid-work>

The Sphere Project. (2011). *The Sphere Project*. Retrieved from <http://www.sphereproject.org/resources/>

The World Bank. (2014). Literacy rate, adult total (% of people ages 15 and up). *World Development Indicators*. Retrieved July 11, 2014, from <http://data.worldbank.org/indicator/SE.ADT.LITR.ZS/countries>

Tisne, M. (2010). *Transparency, participation and accountability: Definitions*. Unpublished background note for transparency and accountability initiative

Tong, J. (2004). Questionable accountability: MSF and Sphere in 2003. *Disasters*, 28(2), 176–89. doi:10.1111/j.0361-3666.2004.00251.x

Torres, R. T. (2005). Reporting. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 371–376). Thousand Oaks, CA: Sage Publications, Inc.
doi:<http://dx.doi.org.ezproxy.library.ubc.ca/10.4135/9781412950558.n484>

Twersky, F., & Arbreton, A. (2014). *Benchmarks for spending on evaluation so everyone can succeed*. Retrieved from http://www.hewlett.org/sites/default/files/Benchmarks%20for%20Spending%20on%20Evaluation_2014.pdf

USAID. (2013). *Technical note: Impact evaluations* (pp. 1–13). Retrieved from http://www.usaid.gov/sites/default/files/documents/1870/IE_Technical_Note_2013_0903_FINAL.pdf

Van Maanen, J. (1988). *Tales of the field: On writing ethnography* (1st ed.). Chicago: Chicago University Press.

Van Wassenhove, L. N., Tomanisi, R. M., & Stapleton, O. (2008). *Corporate responses to humanitarian disasters. The mutual benefits of private-humanitarian cooperation*. Retrieved from <http://www.insead.edu/facultyresearch/centres/isic/humanitarian/documents/R-1415-08-WG.pdf>

Von Zweck, C., Paterson, M., & Pentland, W. (2008). The use of hermeneutics in a mixed-method design. *The Quantitative Report*, 13(1), 116–134. Retrieved from <http://files.eric.ed.gov/fulltext/EJ800290.pdf>

Walker, P. (2005). Cracking the code: the genesis, use and future of the Code of Conduct. *Disasters*, 29(4), 323–36. doi:10.1111/j.0361-3666.2005.00295.x

Walker, P., & Maxwell, D. G. (2009). *Shaping the humanitarian world*. Abingdon: Routledge.

Weiss, C. H. (1972). *Evaluation Research*. Englewood Cliffs, NJ: Prentice-Hall.

Weiss, T. G. (2013). *Humanitarian Business*. Cambridge, UK: Polity Press.

Wenar, L. (2006). Accountability in international development aid. *Ethics and International Affairs*, 20(1), 1–23.

Appendices

Appendix A: List of Evaluation Reports Included in the Research Sample

Barnes, J., Simmons, A., Abola, C., Nkawanzi, V., Semay, S., & Greeley, S. (2013). *Norwegian Refugee Council recovery of acholi youth (RAY) Northern Uganda: Independent project evaluation final report*. Retrieved from <http://www.alnap.org/resource/9950>

Featherstone, A. (2012). *Evaluation of Oxfam's South Sudan humanitarian response*. Retrieved from <http://www.alnap.org/resource/11873>

Duncalf, J. (2013). *Real-Time evaluation of ACF International's response to Typhoon Haiyan/Yolanda in the Philippines*. Retrieved from <http://www.alnap.org/resource/10892>

Gubbels, P., & Bousquet, C. (2013). *Independent evaluation of CARE's response to the 2011–2012 Sahel humanitarian crisis*. Retrieved from <http://www.alnap.org/resource/9703>

HERALD Consultants. (2012). *Final evaluation report of the wet feeding and cash transfer project in Southern Somalia*. Retrieved from <http://www.alnap.org/resource/7771>

Karlsen, M. (2012). *Save the Children emergency response to the Ivorian refugee crisis in Liberia: Final evaluation report*. Retrieved from <http://www.alnap.org/resource/6369>

Kugu, B., & Oksak, E. (2013). *Evaluation study Support to Life's response to Syrian refugee crisis, 2012 & 2013*. Retrieved from <http://www.alnap.org/resource/10304>

Macauslan, I., & Phelps, L. (2012). *Oxfam GB emergency food security and livelihoods urban programme evaluation final report*. Retrieved from <http://www.alnap.org/resource/7074>

Morgan, N., Naz, S., & Sanderson, D. (2013). *Pakistan floods response 2010-12 Evaluation for Christian Aid and its partners*. Retrieved from <http://www.alnap.org/resource/8819>

Shah, R. (2013). *Eye to the future II: Final external evaluation report*. Retrieved from
<http://www.alnap.org/resource/10244>