Georgia State University
ScholarWorks @ Georgia State University

Public Management and Policy Dissertations | Department of Public Management and Policy

4-16-2008

Evaluation of Emergency Response: Humanitarian Aid Agencies and Evaluation Influence

Monica LaBelle Oliver
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/pmap_diss

Recommended Citation
Oliver, Monica LaBelle, "Evaluation of Emergency Response: Humanitarian Aid Agencies and Evaluation Influence." Dissertation, Georgia State University, 2008.
http://scholarworks.gsu.edu/pmap_diss/23

This Dissertation is brought to you for free and open access by the Department of Public Management and Policy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Public Management and Policy Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

EVALUATION OF EMERGENCY RESPONSE: HUMANITARIAN AID AGENCIES AND EVALUATION INFLUENCE

A Dissertation
Presented to
The Academic Faculty

By

Monica LaBelle Oliver

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Public Policy

Georgia Institute of Technology
Georgia State University
August 2008

Copyright © Monica LaBelle Oliver 2008

Evaluation for Emergency Response:
Humanitarian Aid Agencies and Evaluation Use

Approved by:

Dr. William L. Waugh,
Adviser
Andrew Young School of
Policy Studies
Georgia State University

Dr. Gordon Kingsley
School of Public Policy
Georgia Institute of
Technology

Dr. Gary T. Henry
Department of Public Policy
The University of North
Carolina at Chapel Hill

Dr. Kent Glenzer
Institutional Management
and Learning Team
CARE

Dr. Kim Reimann
Department of Political
Science
Georgia State University

Date Approved: April 16,
2008

Evaluation is creation: hear it, you creators! Evaluating is itself the most valuable treasure of all that we value. It is only through evaluation that value exists: and without evaluation the nut of existence would be hollow.

--Friedrich Nietzsche, Thus Spoke Zarathustra

ACKNOWLEDGEMENTS

I am enormously grateful to my husband, Erik, for his unwavering support of me. His willingness to take on the lion's share of the breadwinning burden in our household during my years of study, to say nothing of his cheering for each little accomplishment along the way and editing the occasional page, has made this whole journey possible and bearable. Wonderful Lucas, my 18-month-old ray of sunshine, has been the bright light during my many hours of forcing my extraverted self to sit alone staring foggily at a computer screen. My parents, brothers, and sisters have offered me essential encouragement; thanks especially to Mom for helping me to identify what I am really trying to say in my writing for 20+ years now. I thank LaTrelle for the many hours of help with Lucas' care when I would otherwise have been all too willing to let him distract me. I am grateful to CARE for allowing me to peek inside their operations and to review documents and interview personnel. Finally, I thank my committee members, who have each taken a great deal of time to help me grow as a researcher and conceive this project.

TABLE OF CONTENTS

Acknowledgements iv

List of Tables vi

List of Figures vii

Chapter I, Introduction 1

Chapter II, Review of the Literature 32

Chapter III, Methodology 63

Chapter IV, Findings 90

Chapter V, Discussion and Conclusion 134

Appendix I, List of Evaluations Reviewed 159

Appendix II, Matrix of Lessons-Learned and Evaluations .. 163

Appendix III, CARE Organigram 186

Appendix IV, Interview Protocol 187

References 191

Vita 196

LIST OF TABLES

Coding Scheme for Interview Data..... **86**

Instances of Influence..... **131**

LIST OF FIGURES

Mechanisms Through Which Evaluation Produces Influences.. **47**

Four Levels of Influence for Large INGOs..... **120**

Pathway of influence for Tropical Storm Jeanne..... **126**

Pathways of Influence for India Earthquake Evaluation... **127**

SUMMARY

Organizational development is a central purpose of evaluation. Disasters and other emergency situations carry with them significant implications for evaluation, given that they are often unanticipated and involve multiple relief efforts on the part of INGOs, governments and international organizations. Two particularly common reasons for INGOs to evaluate disaster relief efforts are 1) accountability to donors and 2) desire to enhance the organization's response capacity. This thesis endeavors briefly to review the state of the evaluation field for disaster relief so as to reflect on how it needs to go forward. The conclusion is that evaluation of disaster relief efforts is alive and well. Though evaluation for accountability seems fairly straightforward, determining just how the evaluation influences the organization and beyond is not.

Evaluation use has long been a central thread of discussion in evaluation theory, with the richer idea of evaluation influence only recently taking the stage. Evaluation influence takes the notion of evaluation use a few steps further by offering more complex, subtle, and

sometimes unintentional ways that an evaluation might positively better a situation. This study contributes to the very few empirical studies of evaluation influence by looking at one organization in depth and concluding that evaluation does influence in useful ways.

CHAPTER I

INTRODUCTION

1.1 Overview

Evaluation as a form of research is a practical endeavor, manifested in action. Whether evaluation succeeds in ameliorating the practice of the organizations that employ it is a question that is gaining in relevance for non-profits as individuals, corporations, and governments increase their calls for greater accountability. At its very root, the word 'evaluation' means 'to seek out the value of;'(*Merriam-Webster Online Dictionary*, 2007) when an organization evaluates its policies or programs, by definition it is looking to determine value. It is not surprising, then, that the concept of *use*, or the newer and more apt concept of *influence*, receives such enduring attention in the evaluation literature. As an enterprise that purportedly exposes the value of a policy or program, evaluation can generate concern over what its consequences will be.

In the past two decades, both donors and the public have called for higher and higher standards of

accountability in international and domestic nonprofits. Evaluation, consequently, has come under the spotlight as a means to that end. The lingering question is whether the influence of evaluation on an organization is merely "window-dressing", whether it includes or is limited to accountability. Nonprofits, working with limited resources, cannot afford for evaluation to be ineffective. This paper addresses the question of how the evaluation process and evaluation reports affect (or do not affect) the practice in a large well-known non-profit organization and identifies the factors that expose these influences.

Evaluation use, or utilization, has been prevalent in the evaluation literature for a few decades and refers to a change resulting from an evaluation or an evaluation report. Evaluators or organizations commissioning evaluations often place high priority on using findings for program improvement or, more broadly, to inform decision-making. Use has been a central theoretical theme both for evaluation and for research-generated knowledge. However, organizations often do not have formal mechanisms for assessing whether their evaluations are used effectively or not, and virtually none have a systematic means for identifying indirect or unintended consequences of evaluation. In the case of nonprofits, this could be due

in part to limited resources; moreover, those donating to nonprofits might prefer a more "direct" use of their money than meta-evaluation activities.

In the evaluation literature, the term 'use,' and its sister 'utilization,' have given way to the broader 'influence' (Henry & Mark, 2003; Kirkhart, 2000) in more recent theoretical musings on evaluation. This shift reflects the limitations, semantic and otherwise, of the concept of use, and opens the door for unintended and subtle consequences of evaluating within an organization. Few organizations have studied formally the instance(s) of evaluation influence; perhaps this is in part because the factors leading to an evaluation's influence are poorly understood. Henry and Mark (2003) tackle this in examining the mechanisms which, for them, undergird the instances of influence. A better understanding of these mechanisms, and concrete examples, will go a long way toward helping evaluators to dissect how it is that evaluations potentially influence, and under what circumstances. This study contributes toward developing that understanding of the mechanisms leading to evaluation influence by exploring the extent to which an evaluation's influence can be anticipated or planned for prior to the evaluation's being carried out.

The twenty-first century has seen technological advances, cultural trends, political situations, and globalization unimaginable one hundred years ago. In this short time frame, humanitarian response to natural disasters and other emergencies has increased rapidly. Disasters and conflicts themselves, of course, have been occurring for thousands of years. But never has disaster media exposure made more individuals – and donors – aware of the extent of devastation in disasters than in this century. With this added awareness comes an increase in money flow and in response. While greater emergency response is a welcome development, it brings with it a host of new challenges, such as being accountable for responsible use of donor funds, standards for training of crisis respondents, and timing of assistance, communications, and security.

Among the most critical challenges for a humanitarian aid agency, such as an INGO (international non-governmental organization), is the dilemma of how to measure its impact on a disaster scene. This problem becomes murkier still when aid agencies look to use evaluation findings to improve their efforts for responding to the next disaster. This paper examines the influence, intended and otherwise,

of humanitarian relief evaluation findings in practice and in organizational policy-making.

1.2 Research Problem

In view of the 2004 Asian tsunami and other high-profile emergencies such as the present crisis in Darfur, international humanitarian aid agencies offer a rich forum for examining evaluation influence on an agency's practice. Not only are such agencies grappling with how to use an unprecedented amount of aid money responsibly and effectively, but they are in the business of saving lives. Neglecting to employ the recommendations from an evaluation report could literally be a matter of life and death. These agencies have been facing mounting pressure from donors and governments to strengthen their accountability practices and ensure transparency. As a result, humanitarian aid agencies are more and more frequently obliged to conduct an evaluation as part of their programming. The question is whether evaluations go beyond fulfilling donors' accountability documentation requests and affect agency programs and policies.

In the 1990s, a profusion of large-scale humanitarian crises captured the world's attention. Notably, media coverage of the genocide in Rwanda and Burundi awakened the

public to the urgency of the human suffering there. The famine that befell conflict-ridden Somalia similarly required a substantial response from the international humanitarian assistance community. The upheaval in Bosnia exposed a need not just for relief and refugee assistance, but also for rebuilding and development. These events and others sparked large-scale and visible responses from INGOs. Significantly, the money flowing to aid agencies from governments and from private sources meant a greater call for responsibility on the part of the INGOs whose programming depended on those funding venues. Furthermore, INGOs, who often had an established presence in a crisis area prior to the emergency response, began to partner with the United Nations and with bilateral donor agencies for service delivery and coordination. These relationships increased the clamor for evaluation of emergency programs (Wood, 2001). Public concern for INGOs' comportment grew, with a corresponding demand for better accountability (Ebrahim, 2003b). The number of INGOs increased as well during that period. Though a spate of literature existed at that point on evaluation theory and practice, little of it addressed the unique needs and characteristics of the humanitarian aid agency.

Types of donors to humanitarian assistance organizations range from governments to INGOs to individuals. Government money for humanitarian aid is often channeled through government agencies established for this purpose, as with the United States Agency for International Aid (USAID) in the U.S. or the Department for International Development (DFID) in the United Kingdom. Money designated for emergencies or disasters is usually coupled with development aid budgets (Cahill, 2003). Donor money sources include tax dollars, corporate and individual contributions, and in-kind assistance. As the amount of public and private aid has exploded in the last decade, so, too, has the number and variety of INGOs and other agencies working in emergency relief. The United Nations Development Programme (UNDP), the United Nations High Commissioner for Refugees (UNHCR), and the World Food Programme (WFP) collaborate with INGOs to provide coordination and direction to large-scale relief efforts. The International Red Cross is a familiar figure in these scenarios. Upstart INGOs join the pool in seeking funding and publicity with every disaster. However, as of the late 90s, there were eight INGOs vying for and receiving over half the total relief money: the Cooperative for Assistance and Relief Everywhere (CARE), Save the Children, World

Vision International, the Oxford Committee for Famine Relief (Oxfam), Médecins Sans Frontières (MSF, or Doctors Without Borders), European Solidarity Towards Equal Participation of People (Eurostep), Coopération International pour le Développement et le Solidarité (CIDSE), and Association of Protestant Development Organizations in Europe (APDOVE) (Simmons, 1998). If they are receiving relief money from the same sources, they are also receiving similar pressure to hold themselves accountable for their decisions and actions.

Simmons (1998)likens the competition among INGOs vying for funds and media attention to a market system; Smillie and Minear (2004)similarly label it an enterprise. This image of aid agencies as "corporations" in a "market" hints at the potential for competitive interaction between organizations and underscores the importance of transparency and of organizations holding themselves accountable for their "bottom line." INGOs, regardless of common aims such as a desire to reach the greatest number of people as efficiently as possible, compete amongst each other both for funds and for share of the relief spotlight in the media. The perceived commercial character of the INGO sector has provoked criticism in terms of how INGOs operate and what motivates those who work for them

(Dichter, 1999; Fowler, 1997). This criticism further emphasizes the importance of accountability and evaluation for INGOs that face pressure to show how efficient they have been with donated funds.

1.2.1 Disaster-specific Evaluation Standards

The attention to the importance of accountability heightened in the 1990s in tandem with the increased humanitarian relief response to a series of crises, but was by no means a new issue of concern for INGOs. INGOs have long faced criticism for not making accountability a priority; the very word "non-" in non-governmental organization suggests that INGOs are not beholden to governments, nor bound by their regulations, in the way that corporations, government agencies, or even domestic nonprofits might be (I. Smillie, 1997). The nature of INGO work makes it notoriously difficult to evaluate: it is one thing for a corporation to use sales as a benchmark for product success, quite another for an INGO to measure a concept so nebulous as "empowerment" in concrete terms. This is especially true for INGOs working in humanitarian relief, because often the problems they address are not only complex, but also unanticipated; in a crisis situation, obtaining baseline data can be daunting or

downright impossible. The emphasis on accountability resulting from the wave of emergencies and crisis response funding in the 1990s brought to light a need for setting widely applicable minimum standards for emergency response. Five organizations or consortiums have attempted to address this need through developing guidelines both for planning and for evaluation. Their foci range in breadth from crisis identification and preparedness to post-crisis learning and reflection. These initiatives profess to be complementary to each other, rather than competitive, in function:

1.2.1.1 The Red Cross Code of Conduct

In 1994, the International Red Cross and Red Crescent Movement sought the assistance of established INGO networks to draw up a Code of Conduct for those agencies working in emergency relief. The Code of Conduct is designed to recognize the internal and external pressures that INGOs face in responding effectively and responsibly to emergencies. It is not intended as a mechanism through which to sanction those who agree to it but fail to comply with it; rather, it is a professional guideline. Eight of the largest disaster relief agencies signed on to the Code of Conduct at its inception in 1994, and many more have

since; it aims to be applicable to small and large agencies alike. The Code of Conduct is not so much an evaluation measure or tool as it is a statement of an acceptable behavior.

1.2.1.2 The Sphere Project

Sphere, initiated in 1997, is a collaborative effort on the part of several international NGOs and the International Red Cross. The project is predicated on the notion of the human right to dignity. This, for Sphere, translates to a right to assistance for those whom a disaster affects. Its main products are a humanitarian charter, a framework for quality and accountability in humanitarian assistance, and a handbook of tools for assuring quality in four areas of response:

- water sanitation and hygiene
- nutrition and food aid
- shelter
- health.

The handbook, *The Sphere Humanitarian Charter and Minimum Standards in Disaster Response*, aims to improve accountability and quality of service provision. Each standard has key indicators and guidance notes. The handbook was adopted in 1997 and revised in 2004. Sphere

targets the practitioner and is more of an implementation tool than an evaluation guide.

Sphere is noted for taking a "rights-based" approach to disaster response. The rights-based approach, as compared to a needs-based approach, focuses on acknowledging basic human rights as opposed to needs-based service delivery. In other words, rather than approaching disaster response as fundamentally addressing a *need*, such as drinking water, Sphere approaches disaster response as addressing a *human right*, such as the right to adequate food and water. The rights-based approach endures criticism for not getting to the heart of the political context surrounding a crisis (Hilhorst, 2002). Sphere also receives criticism for being so general that it is difficult to adapt to a particular context (O'Donnell, 2002). Sphere has also come under scrutiny for representing the ideas and priorities of developed-world professionals, thus leaving little or belated opportunity for beneficiaries to lend a voice (Dufour, 2004).

1.2.1.3 ALNAP

The Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) is, like Sphere, an inter-agency collaborative effort. In addition to

attempting to facilitate learning and improved accountability, ALNAP serves as a repository for evaluative reports of the relief efforts of various groups. These are intended as a resource for further learning. It publishes an annual Review of Humanitarian Action in which it compiles evaluation learning from member agencies. It also provides them with its "proforma," an evaluation guide.

ALNAP is particularly focused on sharing of knowledge between organizations. It emphasizes improved quality as a continual goal(Hilhorst, 2002). Better quality of information exchange will lead to better tools for future decision-making.

1.2.1.4 People-In-Aid

People-In-Aid is a human resources-oriented project that maintains its Code of Good Practice. The Code names seven main principles and accompanying indicators and focuses on a context-rich social audit approach. It is engineered as a tool for human resources management, but professes to be adaptable to other audiences. It is not designed to be an instrument for measuring the success of a program; it is simply a guideline for practice.

1.2.1.5 Humanitarian Accountability Partnership International

The Humanitarian Accountability Partnership – International (HAP), established in 2003, is a regulatory entity for international humanitarian assistance efforts. Its ultimate mission is to respect the rights of the intended beneficiaries of humanitarian aid through holding aid agencies accountable to those beneficiaries. Consultation and research on the part of HAP has led to its development of seven core principles of accountability. Members of HAP commit to striving to uphold these principles in theory and in practice through self and external regulation.

In accordance with its seven core principles of accountability, HAP has developed accountability standards against which its members can measure their own accountability practice and identify their strengths and gaps. The accountability and quality management standards are not meant to duplicate the HAP core principles; rather, they provide an instrument for verifying whether a given agency upholds a minimum level of accountability to its beneficiaries. HAP has an eventual goal of creating a certification system through which humanitarian aid agencies can seek officially to be named as being in compliance with the core principles of accountability. This is HAP's main distinction from Sphere, the Code of

Conduct, and People-In-Aid: it endeavors to be measurable and to measure.

Both practitioners and academic researchers use these sets of codes and standards not only for planning relief efforts, but for evaluating them. Some commonly cited strengths of these resources include inter-agency collaboration and a rights-based approach. The weakness most often mentioned is the difficulty of complex situational/contextual dimensions to which these fairly generalized standards do not adapt well. Also, the complexity of arranging for beneficiary participation in these relief efforts is an oft-cited challenge for these standards. Both Sphere and ALNAP are committed to a minimum level of transparency and accountability in the humanitarian assistance community. Evaluations have the potential to be a tool for promoting accountability and for learning within an organization. How to assess an evaluation's impact in terms of promoting organizational learning is a difficult question given the variety of types of evaluations and the range of quality within evaluations.

Though theory-driven articles on evaluation use, utilization, and influence abound, there are few accessible examples of tracking that influence within organizations. In part, this has to do with the difficulty in defining

what constitutes influence within a given organization and where that influence might surface. Aside from direct, instrumental use of evaluation findings, instances of evaluation influence are hard to extricate from other influences in organizational operating procedures and policy decisions. Furthermore, evaluation influence is new and complex enough that few structures exist as starting points for an organization wishing to assess the nature and extent of evaluation influence within its practice.

INGOs have faced mounting pressure from funders and from governments to assure their transparency in their appropriate use of funds to conduct humanitarian work. Indeed, compliance and oversight is a main purpose of evaluation: Mark, Henry, and Julnes (2000)list the assessment of how an organization meets rules and expectations among their four purposes for evaluation. Chelimsky (1997)discusses the "accountability perspective" as one of three core perspectives an evaluator takes in approaching an evaluation. Accountability, then, is a central function of evaluation.

It is one thing for organizations to confirm that they comply with the expectations of major donors by conducting an evaluation whose findings point to appropriate use of

donated funds. This is certainly a critical role for evaluations of humanitarian relief efforts. But if evaluation also influences humanitarian aid organizations in other ways, particularly in serving the purpose of program and policy improvement (Mark et al., 2000), the humble evaluation has the potential to play a central role in helping humanitarian assistance INGOs to fulfill their missions of improving, and even saving, lives. The main research problem addressed here is: evaluation findings may serve a purpose beyond accountability in humanitarian assistance organization, specifically one of project, program, and policy improvement. A dearth of recorded examples exists of how and whether evaluations influence humanitarian relief organizations, to say nothing of the ingredients for a successful instance of influence. This research problem points to two main research questions.

1.3 RESEARCH QUESTIONS

A. Do evaluations of INGOs' disaster relief activities go beyond accountability to affect INGOs' practice?

Though program oversight is a primary reason for initiating an evaluation, a report that actually improves on service delivery is valuable in a sector whose mission has to do with bettering lives.

B. If evaluations of disaster relief activities do improve subsequent projects, what elements of the evaluation process or report contribute to the evaluation's influence on programs and policies?

If indeed such evaluations do have an impact on the decision-making around policies and programs, an organization would benefit from understanding the factors that contribute to the evaluation's being employed positively. Those factors may be intentional or unintentional.

1.4 THE CASE OF CARE

To address these two questions, this study considers the case of CARE, whose emergency response work is well-recognized. CARE's humanitarian assistance work is a compelling single-case study because it is both representative of typical INGO work in this area, and unique in its particular structure and dynamic. Virtually all of the evaluations considered in this study were of responses to emergencies to which other prominent INGOs also responded. By analyzing the case of CARE, this research purports to offer a window on the evaluation characteristics and dynamics typical of INGOs with a similar degree of reach and exposure.

CARE got its start in the 1940s directly after the second World War, when it formed to offer succor to war survivors. Though the organization's work today is not limited to humanitarian relief, its main mission is to fight poverty, in part through emergency response. Whether the emergency is a natural disaster or a conflict, CARE participates in several types of projects, including temporary shelter construction, food and water provision, medical care, and rehabilitation strategies such as economic development. CARE's name is so well-recognized that it received a staggering amount of donations following the Asia tsunami of 2004. Because of this name recognition, it is particularly critical that CARE hold itself accountable to itself, its donors, and its beneficiaries. Because the organization dabbles in so many areas of relief and development, including education, HIV/AIDS, economic development and water/sanitation, its ability to follow up on evaluation findings to improve programming can distinguish it as a premier organization in the world of emergency response.

CARE, as one of the eight largest and farthest-reaching agencies involved in humanitarian assistance (Cooley & Ron, 2002), has been a major recipient of the increased flow of funding to disaster relief efforts, both

from USAID and from private individuals and foundations. As a signatory of the Red Cross Code of Conduct and as an active member of Sphere, ALNAP, and HAP, CARE is a forcible presence in the multitude of initiatives to harness humanitarian aid organization accountability. CARE's current emergency response efforts extend from supplying food and water to providing shelter to facilitation of health care provision and delivery of essential supplies.

Though CARE has periodically conducted in-depth evaluations of its emergency response activities, only in recent years has the organization made a concerted effort to evaluate every emergency response project with an eye toward transparency through making evaluation reports widely available. Availability is not limited to internal CARE staff; large evaluations are also available to other NGOs and agencies involved in relief work. This commitment to transparency reflects an international focus on coordinating and improving on current emergency response practices. The choice of CARE as a study subject precludes an overly general conclusion as to evaluation influence in nonprofit organizations. An international INGO with multiple large funding sources is part of a small group of contextually distinct nonprofits for whom milieu is so integral to their operations that their accountability

practices do not mimic those of smaller-scope nonprofits with fewer funding resources(Ebrahim, 2005).

Within CARE, organizational decision-making happens within several tiers. CARE itself is a confederation of 11 member groups, and so its decision-making is not strictly hierarchical. Policy decision-making occurs at the executive level. An executive group meets (twice per year) to examine priorities for the coming year. This group also does strategic planning for the organization on a five-year basis. For program-level decision making, biannual meetings of program directors and officers result in an agenda for the coming fiscal year. The same is true on a country-by-country level. Moreover, a disaster invites ad-hoc planning of the sort necessary for dealing with an unforeseen situation. Many planning and decision-making events occur in this manner as well. Because of its confederation structure, CARE will not be a case study directly applicable to all agencies its size; nevertheless, its evaluation systems and challenges will be relevant to organizations engaged in similar work.

CARE International's emergency response evaluations will serve as the case study for these research questions. An ODI report (Willits-King & Darcy, 2005)on agencies that respond to complex emergencies found that CARE has a much-

deserved reputation among beneficiaries for responding effectively to emergencies, but that the organization invests little in emergency response by comparison to its peers. The report suggested that CARE is recognized for the quality of its work in humanitarian relief, but that the organization could commit more resources and better utilize existing resources to scale up and increase its capacity to respond.

The nonhierarchical structure and the multinational scope of CARE's work present particular challenges for evaluation. It is difficult to limit any evaluation in the organization to one single unit of analysis because the information needs of different bodies in the organization are very different. It can be difficult, as well, to gather consistent data in such an organization because of logistics and language and cultural obstacles. Also, there are multiple relationship dynamics that come into play when considering the areas of influence on which evaluations touch. For these reasons, CARE is an intriguing case study whose patterns of influence have implications for how similar complex transnational organizations function.

The study draws from two main data sources. The first of these data sources is all of the available evaluation

reports from CARE's emergency response activity from 2000-2005. These reports range from a brief summary to a multi-document behemoth. CARE uses four main formats for evaluation of emergency response efforts. 1) The Real Time Evaluation occurs in the middle of an intervention and assesses the success of the effort so far. 2) The After Action Review occurs just after an intervention and is typically a reflection session lasting three or four days and involving the staff members, temporary and permanent, who comprised the emergency response team. The 3) Final Evaluation occurs after the intervention and formally formulates the lessons CARE hopes to take away from the experience of the response for the future. 4) The Multi Agency Evaluation involves the major INGOs who collaborated to mount a response in a large-scale emergency. These evaluation assess not only the effectiveness of each respective INGO, but examine the collaboration and coordination among all of the INGOs.

The second source of data is a series of interviews with 25 different people associated with the evaluation process, from evaluators to field workers to management team executives. These individuals offer insight into their own perceptions and experience of whether and how the evaluations influence CARE and their practice. The

interviews examine both the perspectives of those who conduct the evaluations and the perceptions of those who purportedly read and use the final evaluation reports.

The research draws upon Kirkhart's (2000)Integrated Theory of Influence, which considers evaluations using three different gauges: Intention (intended or unintended), Source (process or results), and Time (immediate, end-of-cycle, or long-term). These dimensions inform the interview questions around whether and how the evaluations from 2000-2005 affected later practice and policy.

The central analysis for the evaluation and interview data employs Henry and Mark's (2003)'pathways' of evaluation influence as the basis for examining how an evaluation affects an INGO from start to finish. Like Kirkhart, Henry and Mark find 'use' to be a limiting term, and they advocate for the broader 'influence.' Their work culls from social science theories to propose pathways of influence which help to categorize the different levels at which influence might occur: 1) the individual, 2) the interpersonal, or 3) the collective.

Henry and Mark's taxonomy offers a starting point for examining one organization's treatment of evaluations in the emergency response arena. The analysis will use these

three levels of influence to track interview responses and evaluation report data in order to observe how the report process and findings influence various levels of communication, if at all.

Their taxonomy proposes a sort of menu for identifying and categorizing examples of influence. They are drawing from multiple disciplines, so some of their influence categories are more likely to show up in a large, decentralized INGO such as CARE than are others. For example, in the "individual" level of influence, the "attitude change" mechanism is likely to surface in a study of CARE because it is a mechanism that easily lends itself to a program (as opposed to a policy). Determining whether an individual's attitude shifted is entirely feasible with interview data. Conversely, the "salience" mechanism is more about policy-related issues than about programs, and so is not as likely to emerge from the CARE study. Other mechanisms, such as "elaboration," are difficult to pinpoint with interview data. Finally, it is more likely that the study will reveal examples of the individual-level mechanisms and the interpersonal-level mechanisms than the collective-level mechanisms. This is because it is easier and takes less time to effect change at a programmatic level than at a policy level. Moreover, one of the five

pre-interviewees for the study mentioned his own impression that evaluation reports stop short of having policy-level influence at CARE, in part because the culture of learning there does not leave room for evaluation data in executive team agendas.

1.5 DEFINITION OF TERMS

The following terms will be useful for understanding this study:

INGO – international non-governmental organization. For the purposes of this study, the acronym 'INGO' describes any internationally-operating not-for-profit organization or agency.

Humanitarian aid – this work is distinct from development work and the term refers to outside assistance for communities suffering from urgent crises such as natural disasters, conflicts, droughts, or famines.

Use and *utilization* – these terms will refer to the employment of evaluation findings in decision-making. The review of the literature and the discussion of methodology will further detail these and will also distinguish between evaluation use and evaluation influence.

Lessons learned – though this term seems at times synonymous with 'recommendations' in CARE's evaluation

reports, for the purposes of this paper it will refer to those items identified through an evaluation that point to room for improvement and that bear consideration in planning for future emergencies.

1.6 DELIMITATIONS AND LIMITATIONS

This study looks specifically at CARE International's use of evaluation reports for its emergency response projects. CARE, as one of the largest NGOs working in emergency relief, is a good case for the assessment of how evaluations influence the organization at different levels.

This study was limited by the geographic location of the researcher. Though CARE is a confederation whose secretariat is based in Geneva, Switzerland, CARE USA, based in Atlanta, GA, typically plays a somewhat larger role than the other CARE countries in responding to humanitarian crises. This, combined with the fact that the researcher is based near Atlanta, GA, means that the study contains a disproportionate number of Atlanta-based interviews. In addition, the design of the study, while in-depth enough to provide a good picture of evaluation implementation within the humanitarian aid section of CARE, is not broad enough to treat any other INGO with such

depth. The study sacrifices generalize-ability in the interest of internal validity in treating the case of CARE.

The study will have relevance for all humanitarian aid agencies working on an international level, but will be limited to an in-depth study of CARE International's emergency response evaluations from 2000-2005. The two-part study includes a meta-evaluation of all evaluation reports conducted within or on behalf of CARE for its humanitarian aid activities from 2000-2005. This meta-evaluation will be coupled with interviews with 25 CARE employees and consultants who are involved in the evaluations whether as evaluators, as field workers, as senior management, or as middle management.

1.7 SIGNIFICANCE OF STUDY

Existing research treats the subject of evaluation utilization thoroughly. Though much has been written theoretically about evaluation use, utilization, and influence, few empirical studies have attempted to track evaluation report findings from completion to utilization. Fewer still have looked at evaluation implementation as it pertains to humanitarian aid agencies. In the interest of responding to the recent calls for greater accountability on the part of humanitarian aid agencies, this study aims to address the gap in the literature pertaining to their

implementation of evaluation findings. Improved practice is one of the important goals of evaluation, and improved practice in emergency response can save lives.

1.8 RESEARCHER'S PERSPECTIVE

As a former Peace Corps volunteer, the researcher has the tendency to suspect that INGOs, especially those of CARE's reach and magnitude, are inefficient, bureaucratic, and, at times, too "corporate." Though she was careful to try to lay aside any such biases in conducting interviews, she risks unwittingly coloring her conclusions with preconceived images. The researcher worked with CARE as an intern prior to conducting the research for this project, and as such considers her relationships there to be friendly rather than impartial. She made an effort to have other eyes look at the research material when she thought there was the possibility of missing something due to her being sympathetic to CARE.

1.9 KEY ASSUMPTIONS

A major challenge for this study is isolating different forms of influence for evaluations. Determining whether or not field offices have followed up on operational recommendations, a direct, instrumental form of influence, is straightforward; discerning whether an evaluation affected the social fabric of the organization

is not. That is to say that it is one thing to identify instances of change, but another to be able to attribute that change to the evaluation process or findings. Nevertheless, the exercise of identifying mechanisms of influence is valuable because it illuminates the possible forms that the evaluation influence might take, thus allowing for taking steps toward developing a theory for how evaluations work in the humanitarian aid agency setting.

The study assumes that the designated interviewees are indeed an adequately representative cross-section of those who would learn directly or indirectly from an evaluation in the organization. It is possible that a category of people has been left out, or that the interviews are heavy in one area of representation and light in another. One way of curbing this has been having the design, monitoring and evaluation coordinator for CARE look over the list of interviewees for balance and representation. Another has been to ask the interviewees themselves whom they would suggest talking to, and checking those contributions against the list of interviewees.

Building theory about organizational learning is beyond the scope of this research; the overarching goal is to provide CARE with insights on how to make its evaluation

process more effective for learning and to contribute to the dialogue on how evaluation influence works. It is likely that the results will not generalize easily to other aid organizations. Dissemination of the findings of this research to CARE will be of particular importance, as adding another long report to CARE employees' reading lists would hardly send the right message about evaluation report efficiency and effectiveness.

Evaluation has the potential to help an organization fulfill its mission by complementing its quest for higher quality and greater knowledge. The challenge becomes determining whether the good intentions for the evaluation report's end use come to fruition. Exploring how evaluation eventually influences an organization is a practical step for an organization whose success in responding to a crisis depends partly on its self-understanding of its strengths and resources. If a better comprehension of how evaluation influences the organization and beyond can lead to designing evaluations that have greater influence, personnel whose mission is to mitigate poverty, to alleviate suffering, will be better equipped to do so.

CHAPTER II

REVIEW OF THE LITERATURE

The following review of the literature lays out the case for the importance and relevance of looking at how evaluation influence works in a transnational organization. Beginning with the literature on accountability, the chapter looks at how accountability relates to evaluation and why heightened calls for better accountability necessitate more utile evaluations. A review of how the literature on evaluation use has evolved into a dialogue about the many nuances of evaluation influence exposes the need for more empirical study of how influence works. The review then presents the two theoretical models that inform this study. Finally, a look at the literature on learning organizations provides a point of departure for discussion and reflection on the relationship between the organization's learning culture and the influence of evaluations within it.

The search for literature related to the question of how evaluation reports are used or under-used in humanitarian aid organizations revealed an impressive range of different kinds of evaluation activities in the INGO world as well as a spate of literature on the utilization

of evaluation. In addition to Google Scholar, the researcher searched the following databases for relevant literature: ABI/Inform, Academic Search Premier, Jstor, PAIS, and the Web of Science. She employed the following terms for the search: evaluation use, evaluation utilization, evaluation influence, evaluation and accountability, humanitarian aid and accountability, humanitarian aid and evaluation, evaluation and organizational learning, INGO and learning organization. Her intent with these last two search terms was to ascertain whether others have studied or written about accountability or utilization as part of or as resulting from a culture of learning within an organization.

2.1 Accountability versus Evaluation

The terms 'accountability' and 'evaluation' are often used interchangeably in disaster relief parlance. There is, however, an important distinction between the two. Evaluation purports to add to or improve an organization's accountability. This makes accountability an important piece of the discussion on the influence of evaluation in humanitarian relief, as recent years have seen increased attention to the importance of accountability in such agencies as media and public scrutiny of them rises. Accountability usually refers to documenting how donor

funds were spent or, in the case of government dollars, being accountable to taxpayers for the use of their tax money. Accountability is larger than budget records and is also a statement of who is responsible for what or who is in control (I. Smillie, and Larry Minear, 2004).

Evaluation, on the other hand, subsumes accountability; in fact, the best-known set of standards for program evaluation in the United States is that of the Joint Committee on Standards for Educational Evaluation (*The Program Evaluation Standards*, 1994). These standards aim to guide process and outcome evaluations of programs. The standards fall under four headings: utility, feasibility, propriety, and accuracy. Accountability falls in the propriety category. Mark, Henry, and Julnes (2000) list accountability (their term is 'compliance and oversight' as one of four main purposes of evaluation. This is germane to this study because if evaluations are likely to be initiated within the INGO because of increased concern for greater accountability, the question arises as to whether they can also be used to greater effect in other ways. Ebrahim (2003a)distinguishes between internal and external accountability within NGOs. External accountability refers to how the INGO answers to donors. Internal accountability is the agency's responsibility to

itself to perform well. Alistair Hallam (1998)insists that evaluations of complex emergencies should stress either accountability or lesson-learning, and if they stress lesson-learning, they ought to be participatory. Donahue and Tuohy (2007)contend that in disaster response, "lessons-learned" is a 'misnomer.' Their exploratory study concludes that, in fact, lessons are too often *not* learned. Donahue and Tuohy, through a series of focus group interviews, found that even if lessons are identified, an organization often has few systems in place for institutionalizing the learning.

Kirkby et al. list accountability, knowledge, and development as three perspectives on disaster evaluation (Kirkby, Howorth, Keefe, & Collins, 2001). They emphasize that evaluations have purpose, notably knowledge increase, beyond accounting to donors. They mention the value of inter-agency evaluation for enhanced learning; the Joint Evaluation of Emergency Assistance to Rwanda, they observe, has become a benchmark of sorts for this kind of inter-organizational learning. Smillie (1997), in discussing accountability within an examination on what it means for an INGO to be 'transparent,' asks, "greater than what?" as in, NGOs are always called to greater transparency. But greater than what? The accountability question is the same:

NGOs are called to be more accountable. But more accountable than whom? Organizations such as HAP are answering that question with a set of standards for NGOs to use as a reference point. It is clear that there is attention in the literature to establishing a *minimum* standard of accountability. This begs the question of whether there is a limit to how much effort an organization should put into striving for greater accountability. There is a question of whether there a point at which the costs of evaluating exceed the benefits to the organization of the evaluation outcomes. Of course, it seems lofty and ideal for an organization to strive for ever-greater transparency and accountability. But nonprofits by definition are working with limited resources, and at some point spending resources on accountability practices takes resources away from the programs themselves.

2.2 Utility of evaluation

Though accountability receives a great deal of attention as a main purpose of post-crisis evaluation, utility of the evaluation is of particular importance for assessing response to conflict or disaster situations. There are a number of different 'uses' identified throughout the literature, from a distinction between 'direct' and 'indirect' use (Scriven, 1991),to a

categorization of use as conceptual; symbolic; enlightenment; imposed (Weiss, Murphy-Graham, & Birkeland, 2005).'Intended use' of an evaluation is always a significant consideration, but in an emergency, planning for response can be ad-hoc in nature and can tend to draw from available, rather than optimal, resources. Consequently, organizational processing of positive and negative lessons learned from disaster response can be of particular value for anticipating and preparing for future crises. This research will employ Karen Kirkhart's (2000) "re-conceptualized" theory of use to examine the characteristics of emergency response evaluations along axes of source, intent, and time, as a step toward understanding the evaluation report contents and where those report contents might hold influence. That report data informs the coding of the interview data, for which Henry and Mark's (2003) framework provides the basis.

Current thinking on evaluation use stems from a body of work on knowledge and research utilization that sprang from a governmental focus on social betterment in the 1960s and 1970s. Lyndon B Johnson's War on Poverty brought with it an increased spending, not just on social programs, but also on social science research to inform the poverty battle. With the onslaught of social science research came

a heightened concern for accountability, and it is out of this period that program evaluation became an established area of practice. Technology transfer, also, emerged as a critical research area, with obvious implications for how the research was disseminated.

That research theoretically results in new knowledge was not a novel concept during this period of establishing program evaluation; of greater concern was how practitioners utilized that knowledge. Social science research provided a means of identifying cause-and-effect relationships between programming and results, which was of keen interest to decision-makers (Weiss, 1977). Nevertheless, the factors contributing to effective utilization of research remained to be identified. Much of the scholarly thinking during this period looked at the use of research on a national or policy level; the organizational level was still to come. This attention to accountability did not extend to NGOs during this time. NGOs received relatively little public and private funding (and thus little pressure to measure the impact they were having) until the 1980s (Edwards & Hulme, 1996). The 1990s brought still another increase in funding, and a tandem focus on accountability.

The 1980s saw a shift in scholarship toward attempting deliberately to predict or to influence evaluation use. The need to distinguish between different kinds of use became apparent. Carol Weiss (1977) introduced instrumental use as a term for the classic linear form of use, or the use, perhaps, that the researcher intended: the researcher or evaluator proffers knowledge directly to the user, who in turn uses it immediately and as the researcher envisioned. Conceptual use, for Weiss, is less direct and occurs when a piece of knowledge influences an individual's thinking about a policy or program. Finally, symbolic use occurs when decision-makers use research knowledge for political gain or to justify already-made decisions. Though scholars were distinguishing between different kinds of use at this point, many programs -- including INGO development and relief programs -- were not. ALNAP, formed in 1997, raised the dialogue about evaluation use in the humanitarian assistance world by providing a forum for "lessons learned" jointly and individually. Only in its most recent (Sandison, 2006)research is ALNAP including different kinds of use in its scope.

In the late 1980s, a large contribution to the thinking on evaluation use came in the form of the Weiss-Patton debate (M. Q. Patton, 1988; Shulha & Cousins, 1997;

Smith, 1989; Weiss, 1988; Weiss, 1998). Weiss, as a keynote speaker at the annual meeting of the American Evaluation Association in Boston in 1987, made remarks concerning the role of evaluators in determining the ultimate use of evaluation research. She contended that the information resulting from evaluations is not the only factor that influences decision-making; there are political, even non-rational, characteristics of decision-making contexts that affect how evaluation information is received. Evaluators, then, should focus on good evaluation design and good data. Patton took Weiss to task for this, asserting the primary responsibility of evaluators to those who will ultimately use the evaluation results. His concept of use starts with the evaluator journeying with the stakeholders and helping them identify what they need from the evaluation. The evaluation report, then, will be useful if it provides information that meets those needs. Smith (Smith, 1989)pointed out that both Weiss and Patton contribute important ideas to the concept of use, but that the program *context* is critical to determining who the stakeholders are and how they can or will use evaluation information. Weiss, in addressing the American Evaluation Association again a decade later (Weiss, 1998), acknowledged that she and Patton, in fact, agreed on a number of points,

particularly at the program level. She went on to discuss emerging understandings of what it is that is used. She asserted that use is no longer just about evaluation findings, but extends to influence on other organizations and institutions. This debate paralleled attempts at joint evaluations on the part of humanitarian relief agencies responding to disasters. Multilateral donors, NGOs, and consultants comprised the team of evaluators for a large-scale joint evaluation of the response to the Rwanda genocide in 1994(Wood, 2001). This well-publicized effort showed the importance of context to program effectiveness and revealed the importance of considering the effects of the evaluation findings on multiple agencies and institutions. It is not uncommon to explain the inefficiency in using evaluation findings on the political nature of large organizations (Frerks & Hilhorst, 2002).

In more recent research dialogue, the concept of "process use" has introduced the idea that the very process of evaluating is itself a form of use: an interview can be a type of intervention, data gathering can increase inter-organizational communication, and the evaluator's interaction with stakeholders can provide them with an opportunity for reflection (Shulha & Cousins, 1997). The idea of process use expanded beyond the individual's use to

include organizational learning. Patton's (2002)appealing "intended use by intended users" hearkens to the notion that on some level, use ought to be deliberate. His philosophy that end users who are invested in the evaluation process itself will use the resulting information fruitfully makes the important point that programs are fundamentally made up of human beings, and their involvement has to matter. For some kinds of evaluation, such as empowerment evaluation, the people are the only thing that matter in terms of use, as long as they emerge from the evaluation with the capacity to evaluate themselves (Fetterman 2001).

Patton's 'utilization-focused' evaluation is optimal for ensuring impact on some level. It does not address unintended impact or so called "symbolic use" (Beyer & Trice, 1982). This is not a shortcoming; it is simply to say that Patton's approach is not necessarily the ideal model for all evaluation occasions. He does intimate that 'process use' can facilitate communication (Shulha & Cousins, 1997), a notion echoed by Shulock(2000), who mentions it as a means of 'framing political discourse.' Patton's stance, however, is relevant to some contexts, and not to others; when an evaluation is outcome-oriented, for example, is large-scale, and is not concerned with process

questions, the 'intended users' become hard to work with very personally.

Shulock's concept of use, adding to the language in order to further facilitate discussion of new or revisited knowledge, highlights the importance of reporting results with an eye toward framing future decision-making dialogue. Her argument is further reinforced with her empirical test of her ideas; empirical testing of use is not commonplace and has much potential for demonstrating how one might integrate good theory of use into practice.

The study of evaluation use in an organizational setting led to greater questions about organizational learning and knowledge management. Evaluation utilization relates to the organizational learning environment. This is particularly true if, as Patton (1994) suggests, the process of evaluating *is* the learning environment. This idea of process as a valuable forum independent of a "findings" report is only very recently visible in the humanitarian aid world. There is an emerging distinction – most identifiable in ALNAP's evaluative reports database -- between mid-crisis evaluations, final syntheses, and post-response reviews, which are more about reflection on the part of the response team than about impact measurement.

Conner (1998) posits that in order to progress in our thinking about evaluation use, we must take a different, more macroscopic view; Kirkhart(2000)asserts that language is important when theorizing about evaluation, and that "use" is not a sufficiently precise term for describing the way that organizations employ evaluation findings. She suggests the more expansive "influence," which she contends more accurately captures "…effects that are multidirectional, incremental, un-intentional, and non-instrumental, alongside those that are unidirectional, episodic, intended, and instrumental." Her Integrated Theory of Influence considers evaluations using three different gauges: Intention (intended or unintended, Source (process or results), and Time (immediate, end-of-cycle, or long-term.

The first of her dimensions, *source of influence*, refers to the point at which -- or as a result of which -- the evaluation effects change. This may be at the process level or at the results level. The process level comes from Patton's (1998)(1998) notion of "process use," that is, use that stems from the exercise of evaluating. The idea is that involvement in the steps of evaluating can affect an individual or organization in ways distinct from how the ultimate findings of the evaluation might. For the

purposes of this study, the logic model for CARE's emergency response evaluations includes both the process and the results categories for source of influence: indeed, one format of evaluation in particular (the After Action Review) for CARE lends itself to a "process" source of influence.

Kirkhart's second dimension in her theory of influence is *intention*. Intention in her model is a description of whether and how the result of an evaluation is targeted or purposeful. She distinguishes between manifest intended influence and latent intended influence, that is, articulated influence such as program improvement, and unstated intention such as building a program's credibility. "Unintended" influence is simply that which those conducting or commissioning the evaluation did not anticipate.

The third element of Kirkhart's framework is *time*, which she categorizes as immediate, end-of-cycle, or long-term. Acknowledging that these are arbitrary categories, she nevertheless points out that just as programs evolve at different points in their life cycle, so does the potential influence of evaluation results on the program. Her "time" dimension also considers whether an instance of influence

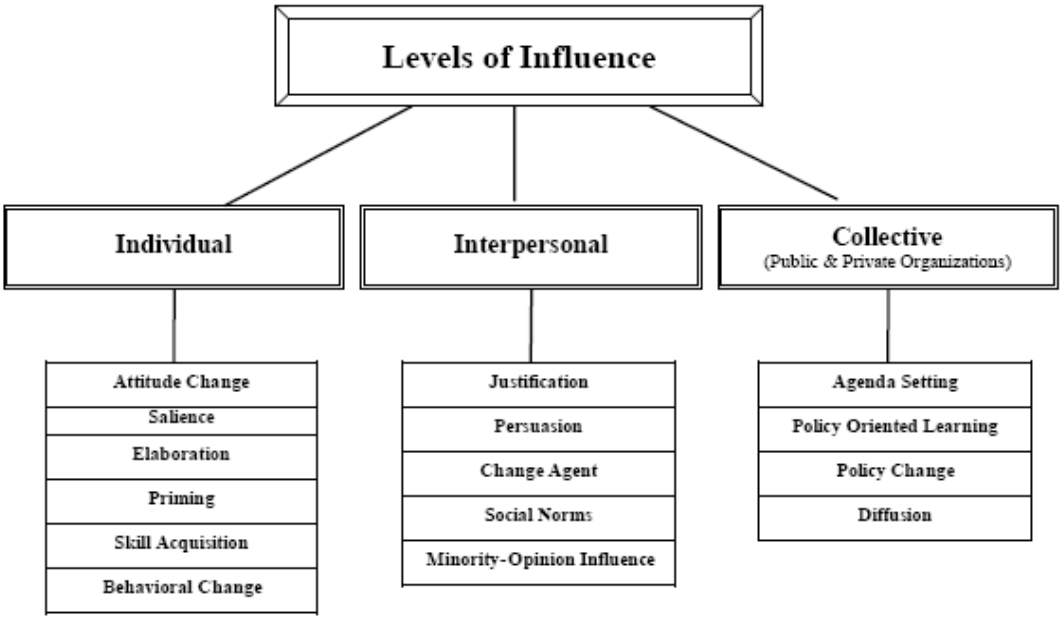
occurs at a single moment in time as opposed to a change process woven through the time period.

Kirkhart's thinking on evaluation influence marked an acknowledgement in the literature that 'use' and 'utilization' are problematic terms for capturing the entirety of the panoply of possible consequences of evaluation results. Henry and Mark(2003)also prefer the broader 'influence,' and offer a framework for representing how evaluation effects various sorts of changes and ultimately leads to 'social betterment.' Their distinction of levels of influence as being between intra- and inter-personal change processes brings up a consideration absent from Kirkhart's three dimensions of source, intention, and time: influence can occur at the level of the individual or at the level of more than one interacting individual. Henry and Mark centrally argue that any evaluation has anticipated outcomes and that mapping influence through the individual, interpersonal, and collective levels can trace change all the way from the evaluation to the policy level.

Henry and Mark's taxonomy, drawing from several bodies of literature in social science disciplines, categorizes evaluation influence into three levels, each of which has several change processes representing what evaluation influence could look like in any given context. Their

levels of influence offer a sort of "menu" from which the evaluator or the researcher may select in order to cater a theory of influence to a particular situation. This figure depicts how their levels of influence break down into levels and "menu" items[1]:

Figure 1: Mechanisms Through Which Evaluation Produces Influences



The individual level concerns change brought about in a single person as a result of participating in an evaluation or reading the findings in an evaluation report. The types of influence for the individual level range from

[1] From Henry, G.T., and Mark, M.M. (2003). *Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions*. Sage: American Journal of Evaluation 24.

attitude change about an issue or program to actual behavioral change. The interpersonal level addresses types of influence occurring between two or more persons, as when one person uses an evaluation's findings to persuade another of his position. The collective level looks at change brought about at the organizational or inter-organizational level, as when an evaluation's findings diffuse to another setting and foster change there.

Henry and Mark's framework proposes a potentially useful tool for categorizing forms of evaluation influence, for tracing "pathways" of influence, and eventually for designing evaluations to have greater positive influence. Very few studies have applied the Henry and Mark framework to a particular case; Weiss et al. (2005) found that Henry and Mark's (2003) framework fit their data well. The results of their study of the effectiveness of the evaluations of the D.A.R.E. program corroborate Henry and Mark's (2003; Mark & Henry, 2004) three change process levels. Their study is particularly interesting because they looked at evaluations of a program widely and publicly considered to be a failure. It would be easy to argue that the evaluations of DARE were irrelevant given the prevailing attitude about the merit of the program. However, Weiss et al.'s application of the Henry and Mark

framework finds that the framework fits the DARE evaluation experience. That the Henry and Mark general framework adapts to fit a local context suggests that evaluation influence is, in itself, a viable concept, and that evaluation itself can cause change in even cause multi-level positive change even in a failed program. The authors also identify a new concept of use – imposed use – where (federal) donors may dictate what an agency does with its evaluation results in order to continue to be funded. This may be relevant to the case at hand; Weiss et al. speculate that a 'results' orientation will foster imposed use, and CARE professes itself to be results-oriented.

Kirkhart and Mark and Henry make good cases for the potential applications of a sound theory of influence adaptable to a given context. Both offer dimensions (Kirkhart's term) that might undergird such a model of influence. Other possibilities for dimensions show up in the literature. Almeida and Báscolo (2006), in their review of the literature on use of research results in decision-making, present the interaction between the researchers and the decision-makers as a potential root of knowledge transfer. This could be a subset of Kirkhart's source of influence dimension. Leviton (2003)finds both the Kirkhart model and the Mark and Henry framework to be practitioner-

friendly distillations of theoretical advances on the topic of evaluation use, though she suggests that there is even still a need to consider context and user knowledge construction. Leviton's supposition is particularly relevant to the CARE example because there are several different learning and operating contexts within the same organization, all of which draw from the same evaluation reports.

Ginsberg and Rhett(2003)look at a series of education evaluations in an attempt to pinpoint where evaluations have influenced legislation and policymaking. Though the thrust of their work is to make a case for scientifically rigorous methodology in response to increasing call for good evidence in the policy arena, they also convincingly observe that not all good questions are causal and that sound methodology can address implementation, not just cause. Their work is an important example of how evaluation influence at a policy level can be traced back to the evaluation itself. A large organization could, perhaps, trace evaluation influence to determine whether the evaluation reports have any influence on organizational policy.

2.3 LEARNING ORGANIZATIONS

Examining the influence of evaluation on a large humanitarian relief agency would not be complete without considering the organization's internal culture of learning. The employees' perception of the culture of learning also surely has an effect on their attitude toward the utility of evaluation. The body of literature on organizational learning is vast and spans decades. Argyris and Schon's seminal work on organizational learning (Argyris & Schon, 1978; 1996)fleshes out the seeming paradox embedded in the concept: individual people learn, retain information, transfer information, and so forth, so how can an organization be said to "learn"? What and how does it learn? They get around this paradox by examining what it is to be an "organization." An organization has procedures and boundaries, and significantly, it designates individuals to make decisions for the whole. If individuals can act on behalf of an organization, then they can learn on behalf of an organization (Argyris & Schön, 1996). They distinguish between different types of learning: *single-loop* learning changes either individual assumptions behind organizational strategy, or changes the organization's strategy. *Double-loop* learning, on the other hand, changes values in addition to strategies and

assumptions. This distinction suggests that there are different levels of learning and that each level depends greatly on the learning environment.

Peter Senge (1992)brought the term "learning organization" into mainstream usage and offered the characteristics of such an organization. Senge named five "disciplines" of a learning organization: 1) personal mastery of individual vision and of objective reality; 2) mental models, or assumptions affecting how we see the world; 3) building a shared vision of the future; 4) team learning and dialogue, and 5) systems thinking, or being able to "see the organization's patterns as a whole…from within the organization." This fifth discipline incorporates the other four and is critical to the organization's evolving as a learning organization with a learning culture. For Senge, managers must learn to strategize, not merely within the scope of their own responsibilities, but about the whole system and in the long term.

Organizational culture may be the single greatest factor in how an organization learns. It is difficult to pinpoint the characteristics of an organizational culture. Certainly, beliefs, values, and norms about performance management are likely to comprise a portion of the

organizational culture, especially for larger organizations. Sackmann (1992) turns to the concept of sense-making to explain what is at the heart of an organizational culture: Individuals make sense of what they experience through cognition. Cognitive aids might include labeling things or events, attaching causes to events, or developing lessons learned to take away from events. When these cognitive aids are common to a group, they comprise some of the collective knowledge that guides the behavior and thinking of the group. Sackmann distinguished between four types of cultural knowledge. *Dictionary* knowledge takes the form of definitions and labels. *Directory* knowledge identifies cause and effect relationships and establishes how things happen or are done. *Recipe* knowledge describes the cause-and-effect relationships of possible events, or how things *should* be done. Finally, *axiomatic* knowledge involves core beliefs, or the "why" behind how things are done. These distinctions of culture and knowledge are relevant to the case of CARE because the organization is so layered and complex that those who receive evaluation reports and are expected act on the findings and recommendations may have information and knowledge needs that are worlds apart. A consideration of what information looks like for the

respective audiences could make the difference in how CARE makes use of the information.

1.4 PRIOR RESEARCH

Humanitarian aid organizations have characteristics that distinguish them from corporations and from other nonprofits in terms of learning environment. Working in a developing country often means confronting an unpredictable, chaotic setting with little infrastructure. Moreover, cultures and levels of development can differ so much from one to another that it appears impossible simply to apply a program to one place just because it worked in another (Berg, 2000). Humanitarian aid organizations often have country-level offices as well as headquarters with divisions across sectors. Each of these layers has a structure based on the competencies and responsibilities of the individual employees. This can encumber learning.

Research on learning in humanitarian aid organizations is as varied as are the organizations themselves. Agencies can be local, working uniquely on emergencies and only in one country. They can be large, as is the case with CARE and its contemporaries such as Oxfam and World Vision. These organizations often work in other areas in addition to emergencies and work in several regions in the world.

There are also governmental organizations and international organizations such as the United Nations.

One study of the UNHCR looked at the UNHCR response to two different crises: the 1991 conflict in northern Iraq, which followed the gulf war, and the Kosovo refugee situation of 1999. The objective of the study was to determine whether any improvements in the UNHCR's response in Kosovo resulted from "lessons-learned" in the Iraq evaluation. The basic finding was that improvements in response did result from the evaluation lessons-learned, but that some recommendations were easier to incorporate than others. For example, recommendations regarding a simple technical operation were easy to implement, whereas recommendations about early warning, which depends on political and external factors, were not.

Another study, of the International Institute for Educational Planning, attempted to answer the question of whether learning takes place primarily through studying the organization's training and capacity-building programs. In another case, ALNAP published a volume of case studies of humanitarian aid agency evaluations (Wood, 2001). Two of the case studies address practitioners' reflections on the follow-up of evaluation recommendations and lessons-learned. One found that designating a formal follow-up

person or team assured that the recommendations would be revisited (Wood, 2001). The other also recommended a formal follow-up structure, and further remarked that the timing of the evaluation report affected whether it was incorporated into planned decision-making meetings. While these studies offer useful insights on which recommendations lend themselves to implementation or how to enhance the process of evaluating, they do not consider evaluation use beyond that which is direct and instrumental.

It is difficult to address evaluation utilization without also considering how or whether an organization learns or has a culture of learning. Ramalingam (2005), in designing a study of several international development agencies, asks: "how does the organization measure the costs and benefits of learning or of not learning? How have systems of monitoring and evaluation been used to map these costs?" Indeed, it is hard to divorce evaluation from an organization's style of learning. This also brings up the question of whether learning is necessarily a priority for NGOs. Evaluation is already an accountability tool, and NGOs are more flush than ever with aid money. There is a question of whether learning in and of itself is an

important enough goal to justify the time and expense of evaluation.

Taut and Alkin (2002) look at program staff impressions of what impedes evaluation implementation. Their study tests whether factors identified by Alkin (1985)as impediments to utilization are the same factors that impinge on effective evaluation implementation. Their study of the UCLA Outreach Staff concluded that, indeed, the barriers to evaluation implementation mimic those of evaluation utilization. Taut and Alkin suggest that the attitudes and perceptions of the members of an organization greatly affect how and whether the evaluation is conducted. Forss, Cracknell, and Samset (1994)also found that the involvement of organization members in the evaluation process will often spur organizational learning.

The recent literature on humanitarian aid reveals that there is an increasing interest on the part of donors, governments, and the public at large in greater INGO accountability. The scandals in the 1990s exposed the dubious practices of prominent NGOs and undermined the public's trust (Gibelman & Gelman, 2001). The intense media scrutiny of these events resulted in an international community suddenly attentive to INGO efficiency and effectiveness. It amounted to a clarion call for NGOs to

hold themselves accountable to a respectable degree of rigor. Ebrahim (2003a) points out that NGOs had theretofore been assumed to have been effective at serving the world's poor, and now must concretely consider how to integrate accountability into their operations.

A cornerstone of an organization's internal system of accountability is its monitoring and evaluation activities. The sequitur from that, beyond merely determining if the organization can adequately and justly account for its actions, is the question of whether and how the organization then uses its evaluations to make positive changes in its practice. Also, the organization may go so far as to employ evaluation findings to better its practice on the ground, but what of the policies that govern the organization as a whole? Ginsberg and Rhett's (2003)experience suggests that evaluations appropriately timed and sufficiently scientific can affect legislative decisions in congress. This leads one to wonder whether evaluations can have a similar pattern of influence within a large-scale organization: is there a policy level of influence distinct from the program level of influence for evaluation results, and are the indicators for success different at the policy level?

SUMMARY

The body of literature on evaluation utilization is substantial, spanning nearly three decades. Much of it focuses on instrumental use, that is, the direct link between research and knowledge transfer. Shulha and Cousins (1997) and others have observed that context is critical to understanding and studying evaluation use. Use is itself an overused and vague term; influence ((Henry & Mark, 2003; Kirkhart, 2000; Weiss et al., 2005) better captures the panoply of impacts an evaluation may have on an organization. Weiss makes the clearest connection between the call for greater accountability and the (imposed) use of the evaluation results. Though her conclusions draw from an American example of a government-funded public program, her observation that a donor's scrutiny of accountability practices can lead to an imposition of how evaluation results are employed going forward is relevant for any sort of agency receiving some public funding.

If dialoguing in terms of evaluation influence, rather than in terms of evaluation use, more accurately reflects how evaluation is a tool for nonprofits to make the world better, the question becomes what exactly "influence" looks like. Kirkhart attempts to answer that question with her

integrated theory, providing a generalist framework depicting influence as a function of time, source of influence, and intent, with subcategories within those. Mark and Henry, also, frame influence, projecting it in levels of communication or interaction. Both of these frameworks, in theory, can adapt to a specific context. This paper will use Kirkhart's framework to develop a context-specific map of CARE's emergency response evaluation influence, informed by Mark and Henry's interpersonal categories. Mark and Henry's model will serve as a springboard for developing a theoretical framework for evaluation in the humanitarian assistance and emergency response context. Kirkhart's model serves as a point of departure for a discussion on evaluation influence writ large; Henry and Mark's framework helps to ascertain whether evaluations go beyond accountability and what factors lead them to do so.

The media frenzy over the scandal-riddled 1990s for NGOs appropriately coincided with emerging scholarship on evaluation influence. Though it is established that there is desire on the part of donors, governments, and the public for greater and more consistent INGO accountability, a larger question remains as to whether the INGO's holding itself more accountable in fact improves on its practice

and furthers its mission. There is a dearth of study on just how NGOs account for their actions and, more importantly, whether their accountability efforts lead to better practice and positive societal change. This study looks at both accountability and evaluation influence in the specific context of CARE's emergency response activities.

The literature supports the theory that there is great interest in INGO accountability and a need for structures that support that as well as better understanding of what it looks like. The evaluation literature shows a progression of thinking on evaluation utilization, with current attention to how to study evaluation influence. But the "so what?" question is whether established accountability systems on the part of the INGO results in better practice and ultimately in the NGOs improving on fulfilling their mission. It is also possible that evaluations influence the organization and beyond in ways, intended or unintended, that transcend the traditional and tangible notion of direct, instrumental use. These less-observable forms of influence may represent valuable ways in which evaluations affect individuals, the organization, or the broader INGO community, whether through participation in the evaluation process or through exposure

to evaluation findings in a report. This study contributes an empirical approach to tracing “pathways” of influence in a large, transnational organization to the literature on evaluation influence and on NGOs.

CHAPTER III

METHODOLOGY

3.1 Introduction

As previously stated, this study aims to examine how humanitarian aid evaluation reports, and the process of conducting them, influence the aid agency's policies and practice. By applying Henry and Mark's mechanisms for influencing to the case of CARE's humanitarian relief work, the study identifies factors that lead (or do not lead) to the application of the evaluation findings and also investigates unintended ways in which the evaluations affect the organization. This study is important because the effectiveness of disaster relief is important for millions of people who are the victims of disaster and evaluation provides one of the best ways for improving policies and practices that can in turn improve the outcomes for those affected by disasters. A deeper understanding of how and why evaluation reports effectively inform and improve practice (or why they do not) can help NGOs to operate more efficiently and with greater accountability to their donors and to their beneficiaries. Evaluation recommendations can improve practice by showing an agency where to change ineffective practice and when to replicate things done well. It is certainly the hope that

evaluations facilitate an agency's being accountable to its donors, beneficiaries, and the public. But it may also be the case that evaluations have a valuable role beyond that for improving policy and practice.

Theories of and Frameworks for Influence

The research specifically looks at how and whether evaluations influence humanitarian aid organizations and the environment in which they operate. Evaluations also affect this environment. Ultimately, the research probes whether this call for increased accountability on the part of NGOs does in fact lead to a positive difference in how they practice. To do this, Kirkhart's Integrated Theory of Influence will serve as an initial frame for sorting out the ways in which 'influence' differs from 'use' in describing what sort of impact an evaluation may have for humanitarian relief. An earlier version of this study proposed to use Kirkhart's theory to describe the patterns of influence at CARE; however, her theory proved difficult to operationalize for this purpose. Her theory is entirely relevant to the discussion on influence versus use and to the broader conversation about the need for useful and accurate theories and models of influence. Her theory is

used here merely to help shape the final analysis by providing points of discussion for what influence looks like in practice

For the specific case of CARE, Henry and Mark's Pathways of Influence serve as the guide for identifying what evaluation influence looks like within CARE's emergency assistance practice. Their framework provides a sort of menu of possible mechanisms through which evaluations may influence an agency and the broader policy community in which it functions. This study will consider each of the menu items for their relevance and applicability to the context of the large transnational humanitarian aid agency. Henry and Mark maintain that culling from social science research yields 'pathways' which help to shed light on the various forms evaluation influence can take. They limit their discussion of evaluation influence to that influence which is relevant to 'social betterment,' an important narrowing of scope given vast number of ways in which 'influence' might manifest itself. Henry and Mark's work is particularly germane to the humanitarian assistance INGO because responding responsibly to emergencies is by its nature a gesture toward social betterment. Since they are likely to conduct the evaluations for accountability reasons alone, such

organizations stand to benefit a great deal from better understanding how their evaluations influence their policies and programs. Further cultivating the aspects of their evaluation process which favor eventual program and policy improvement would amount to better furthering their humanitarian missions.

To identify factors leading to the effective utilization of evaluation processes and reports in humanitarian relief agencies, CARE International's emergency response division serves as an extensive case study of evaluation impact. CARE is a significant case because it is a major deliverer of relief and has a high profile for this type of work. Because of its non-hierarchical organizational structure and its presence in multiple locations, it does not mirror the structure of other INGOs, and it is in a way a series of cases within one. That is, CARE is so decentralized that the regional and country offices act with great autonomy and, therefore, are somewhat independent in the extent to which they are influenced by evaluation. It is nevertheless a good case to study for these questions, as it is a central player in the humanitarian assistance world, and because looking at several layers within the one organization offers a good opportunity for discovering instances of influence at more

than one level. There are other similarly de centralized INGOs (e.g. Oxfam) who presumably would have similar evaluation challenges and patterns. Within the multi-tiered organization, there are sometimes simultaneous relief efforts occurring. The professional staff brings a wealth of expertise to its emergency response efforts, from procurement experts to security personnel to advocacy representatives to evaluators. In contrast to many of its peers, the organization has commissioned thorough evaluations for each and every one of its emergency response activities for the past seven years, with the larger-scale relief efforts evaluated more than once in different formats. This offers a host of evaluation reports for study and comparison.

3.2 Rationale for Qualitative Design

This study employs a qualitative design. The research is based on (1) a series of interviews of individuals who conduct the evaluations or receive/theoretically use the evaluation findings as well as (2) a content analysis of current evaluation reports. Though a quantitative approach would also be possible for this research, such as through a quantitatively-analyzed written survey of those who conduct and receive evaluation reports in the organization, the

choice of a qualitative design is appropriate given the lack of theory in the field. Moreover, while there is a great deal of research on evaluation utilization, very little of it looks at evaluation use using a specific organization or agency as a case, and none of it looks at the international humanitarian assistance context.

The great advantage to using the CARE case to look at evaluation influence is that CARE has a similar approach to emergency response wherever it works, regardless of the type of emergency or the geographical location of the emergency. Some of the same personnel deploy to emergencies in very different contexts. Moreover, the evaluations of each of CARE's emergency response efforts are similarly formatted, with evaluators often involved in assessing more than one emergency. All of this overlap helps to assure that, though the emergencies can vary greatly in terms of type of emergency, scale and setting, there are constants such as the response and evaluation formats that make them comparable within CARE.

This context also means that evaluators are working in a variety of settings, from natural disasters to conflict situations to famines and droughts. The evaluators are preparing reports that have to meet the needs of several audiences, such as for executives or administrators who do

not frequent the field. Consequently, the perception of an evaluation report's utility could vary according to the station or the needs of the individual receiving the report. This difference in perception of quality is far easier to capture in an interview than in a survey or similarly impersonal instrument.

3.3 Research Purpose

The purpose of this research is to identify what factors, if any, enhance or reduce the influence of evaluation processes and reports on practices, programs and policies within a large INGO. The study examines the process leading up to completion of the evaluation report for its contribution to the influence the evaluation has as a whole on the organization. The research will capture the perspectives of the evaluator, the decision-maker who could call for changes based on evaluation reports, and the implementer who would make those changes out in the field. The research can inform CARE, but could be relevant to any INGO working in the humanitarian assistance arena. It should also have relevance for nonprofits in general in that it helps to develop a theory of influence within a specific organization in a particular context. This study has been developed to enhance our understanding of the ways in which evaluations influence humanitarian aid agencies.

The study endeavors to contribute to the development of theory about evaluation influence. Presently, the theories about how evaluation influence works have been laid out, notably by Kirkhart and more thoroughly by Henry and Mark, but little testing has been done on the existing theories so as to confirm or refine them. This study represents one effort to do so.

3.4 Research Questions

This research proposes to address two main questions:

A. Do evaluations of INGOs' disaster relief activities go beyond accountability to affect INGOs' practice?

B. If evaluations of disaster relief activities do improve subsequent projects, what elements of the evaluation process or report contribute to the evaluation's influence on programs and policies?

With these questions serving as an overarching guide, the research applies Henry and Mark's Pathways of Influence to the emergency response context. The framework provides a template for "mapping" the patterns of influence at CARE, steering the discussion toward possible explanations of whether, why and how influence did or did not occur.

Though the study looks at all evaluation reports on humanitarian activity that CARE completed between 2000 and

2005, it examines the example of Tropical Storm Jeanne in greater depth. The reason for highlighting this study is that it provides an example of two different evaluation approaches: an after-action review and a final evaluation. Moreover, in the five preliminary exploratory interviews preceding this study, more than one CARE employee touted Tropical Storm Jeanne as an example of an influential evaluation.

The research draws primarily from two sources of data:

I. CARE has its own repository of evaluation reports from emergency response evaluations conducted over the past five years. A meta-analysis tool (see Appendix) identifies the characteristics of these evaluations. The meta-analysis checklist will help in discerning the patterns in the lessons-learned and evaluation findings that should theoretically be feeding back into the organization's programming and policy-making.

II. A stakeholder analysis will be the basis for the second source of data, a series of interviews with CARE personnel. These interviews target a representative group of persons at various levels of the organization's emergency relief programming.

Interviews focus on evaluation report terms of reference, dissemination, format, timing, and end use.

For the first exercise, the study will look at all of CARE's evaluation reports from its emergency response from 2000-2005. The different categories of evaluations are as follows:

A. Final Evaluation (FE). These long, thorough papers were generated from evaluations conducted after an intervention has concluded.
B. After Action Review (AAR). Also called Lessons-Learned Workshops, these evaluations are typically reflection sessions taking place shortly after an intervention and including many or all of the staff involved in the intervention. The reports stemming from AARs are typically shorter than those of final evaluations, and can be in worksheet form as small groups reflect on their experiences.
C. Real Time Evaluations (RTE). These assessments take place during an intervention and are meant to take stock of progress toward project goals.
D. Multi Agency Evaluations (MAE). These involve several prominent agencies assessing their joint

level of effectiveness in responding to a large-scale emergency, i.e. the Asian tsunami.

A meta-analysis of these evaluations, using Kirkhart's Integrated Theory of Influence as a guiding structure, informed the interviews for the study employing Henry and Mark's Pathways of Influence. Her theory, stemming from the notion that influence is a broader and more accurate term than use for the many direct and indirect consequences of evaluation, names three dimensions of influence that target both the impact of evaluation findings and the unintended results of the evaluation process. To map how and whether influence occurs or is perceived to occur within CARE, the research analyzes the meta-evaluation data and the interview data through the lens of Kirkhart's Integrated Theory of Influence framework and its three dimensions.

The first dimension, the *source of influence*, targets the evaluation characteristic that is the foundation for the evaluation's influence in the organization. The two data sources provide material for identifying which evaluation findings (if any) are being utilized and why. Both the process of evaluating and the corresponding results are the potential sources of influence. The second dimension, the *intention*, looks at the systems (or lack

thereof) in place in the organization for affecting the utilization of evaluation lessons-learned. It also attempts to identify any unintended or unsystematic uses of findings within the organization. The third dimension, *time*, considers the chronology of the evaluations and their eventual uses, taking into account short-term, intermediate, and long-term periods of time. Interview responses are coded into each of Kirkhart's dimensions and sub-dimensions, and the 22 evaluation reports are similarly catalogued.

Using Kirkhart's influence framework as a lens for examining humanitarian response evaluations allows for a structured way of looking at whether her three dimensions account for the factors of distribution and dissemination, format, timing, and decision-making patterns, as well as any other factors that emerge. The extent to which her framework explains what happens in the utilization of emergency response evaluations can be instructive in pinpointing what is unique to this genre of evaluations, and what is universal. Furthermore, her treatment of process use as parallel to (rather than as an afterthought to) results-based use is an interesting test for an organization self-professed to be results-oriented.

Specifically, the analysis consists of reviewing the coded interview and evaluation report data and sorting it into a matrix using Kirkhart's categories:

SOURCE OF INFLUENCE

- Process
- Results

INTENTION

- Intended
- Unintended

TIME

- Immediate
- End-of-Cycle
- Long-term

3.5 Framework and Data

The aim of the meta-analysis of evaluation reports was to catalogue their content and format so as to have a picture of CARE's current evaluation scene. This allowed a subsequent construction of a context-specific theory of influence for CARE's emergency response, using Kirkhart's framework as the underpinning. Appendix B is the meta-evaluation checklist used as a basis for reviewing these 25 evaluations. The checklist combines criteria from the following sources:

1) MEGA Evaluations. Beginning in 2000, CARE hired an independent consultant to conduct a meta-analysis of

its evaluations three times. These MEGAs (Meta-Evaluations of Goal Achievement in CARE Projects) include, but are not limited to, emergency evaluations. Among the recommendations of the MEGAs is that all CARE evaluations contain a "lessons learned" section. This is, in fact, characteristic of CARE's more recent disaster evaluations. But are the lessons really learned? If so, what does the learning look like? If not, what might CARE do to ensure that the lessons are better learned in the future?

2) ALNAP Quality Proforma. In 2001, ALNAP developed a tool designed to help organizations determine the quality of their evaluation reports. The Proforma purports to tap into current internationally recognized best practices.

3) CARE International Evaluation Standards. CARE International has a set of standards to which the organization as a whole adheres, in principle.

The primary objective of this first exercise was to establish the patterns and types of lessons-learned and recommendations in CARE humanitarian aid evaluations.

For the second data source, a preliminary scan of the above-mentioned 22 evaluations resulted in the following interviewee categories:

- senior management;
- external relations;
- finance;
- human resources;
- security;
- procurement and logistics;
- learning and knowledge management;
- country and program directors;
- evaluators.

An intern at CARE also scanned the content of the evaluation reports for patterns. Her observations and conclusions confirmed those of the researcher.

In addition to the scan of the 22 evaluations, the researcher conducted five pre-interviews with individuals representing different pieces in CARE's evaluation process. The researcher identified these five individuals with the help of Jock Baker, CARE International's Coordinator for Monitoring and Evaluation for emergencies. The 30 total interviewees, representing various facets of the emergency response process (and varying levels of authority within the organization), answered questions about evaluation use within the organization. Appendix D is a list of interview questions for the second data source. The interview questions have sub-probes which account for the different individuals' roles within the organization. The researcher conducted all of the interviews personally, though CARE's

Coordinator for Monitoring and Evaluation of emergencies accompanied her for the first three interviews to assure that she was interviewing effectively and taking notes that reflected actual discussion. Rather than collect audio recordings of the interviews, which would have been exceedingly difficult given that several of the interviews with individuals residing overseas took place by telephone, the researcher took handwritten or typed notes during the dialogue and then filled in the notes with detail immediately after the interview. In the evening of the day following an interview, the researcher typed up the notes and e-mailed them to the interviewee so that he or she could review the notes for accuracy and any additional comments. Though the fact that the researcher conducted the interviews personally calls into question the impartiality of the data, she suggests that it also assures consistency given the semi-structured nature of the questioning.

Henry and Mark present a theory of influence for evaluation that specifically hones in on which elements of evaluation affect what they term 'social betterment,' the "…improvement of social conditions." This makes their theory an appropriate vehicle for studying the case of

CARE, an INGO whose primary mission is to serve the poor.

ANALYSIS

Henry and Mark liken an evaluation to an intervention in that an evaluation effects change, for better or for worse. That change might come from the process of evaluating or from the evaluations findings; it might be intentional or unintentional. Henry and Mark propose a theory that is to be recognizable in numerous settings. Their theory includes a number of different forms evaluation influence might take, allowing one to use their framework as a guide to formulating a case-specific theory of influence. This study puts forth a theory for the case of CARE which has implications for similar INGOs and other nonprofits.

Drawing from several social science traditions, Henry and Mark delineate three levels of evaluation influence: individual, interpersonal, and collective. The individual level involves a single person altering his or her thoughts or behavior as a result of participating in an evaluation or being exposed to evaluation findings. The interpersonal level includes the effect of the evaluation on the relationship between individuals. The collective level refers to change occurring in an organization as a result of the process or findings of an evaluation. Henry and Mark

break down each of these levels by means of a taxonomy that lays out the forms of influence specific to each level. It is this taxonomy that serves as the fundamental framework for analysis of whether and how evaluation influence occurs in an emergency response organization.

Analysis of the meta-evaluations and interviews draws from Henry and Mark's discussion of possible pathways of influence and their corresponding framework. Their taxonomy of three levels of influence with corresponding forms of influence offers a sort of checklist for identifying influence instances through the interviews of CARE personnel. The interview data will be aligned with the evaluation reports themselves for an exploration of whether the evaluation influences pinpointed in the interview data are consistent with the lessons-learned and recommendations put forth in the evaluation reports.

The analysis considers each mechanism within each level of influence in Henry and Mark's framework. Specifically:

INDIVIDUAL – there are six potential outcomes of evaluation at this level.

Direction or valence of an attitude change – this research scans each interview for examples of a change in attitude,

positive or negative, about the intervention being evaluated or about the disaster relief program as a whole.

Behavior change – this looks not merely at attitude, but at change in how an individual acts as a direct result of participating in an evaluation or reading an evaluation report.

Salience – this category refers to the importance an individual gives to an idea. In the case of CARE and emergency response, interview data will be culled for whether evaluation findings raised or lowered the priorities of the interviewees vis-à-vis CARE's disaster relief approach and strategies.

Elaboration – this refers to the extent to which a person thinks about or mentally processes a given issue. This is particularly pertinent to the CARE case, as one of their key evaluation formats, the After Action Review, has thoughtful reflection as a central component of the evaluation process.

Priming – this mechanism brings a given idea or concept to the forefront, setting it up to have an impact on judgments or decisions. For the CARE case, priming could have as much to do with the evaluation reports as with the interview data; executive summaries may highlight certain issues and give cursory treatment to others.

Skill acquisition – this refers to increasing one's competence in a skill area via participating in the process of evaluation. For the case of CARE, this mechanism has a natural home with the After Action Review, which typically has small group work and could lead to enhanced collaboration skills.

INTERPERSONAL – the types of influence falling under the interpersonal column include:

Justification – this involves using the conclusions from an evaluation report to back up one's prior convictions about an issue. This can be either a positive or a negative phenomenon. Though an important role for evaluation findings, justification may be difficult to observe through interview data because few if any of the interviewees would have been in a position to use an evaluation report this way.

Persuasion – This refers to attitude change that one individual attempts to bring about in another; evaluation findings can be a central tool here. In the case of CARE, this might be a way in which an evaluation of one response might yield findings that one individual uses to persuade another to act in a certain manner for a future emergency response.

Change agent – this is where participating in an evaluation or reading the findings from an evaluation can lead an individual to take focused action on bringing change about. For the CARE case, this could manifest itself in country directors who help to lead CARE's response for an emergency, and then find themselves in other emergencies later.

Social norms – these are agreed-upon principles about how to conduct oneself in a given setting. This could be difficult to observe through interview data, as change in norms can be so subtle that those involved are not aware of the change taking place.

Minority-opinion influence – This mechanism has to do with altering the opinion of those whose attitude does not align with the majority. Evaluation findings can be a tool for encouraging attitude change.

COLLECTIVE ACTION – there are four versions of influence at this level.

Agenda setting – this is about getting an issue on the docket for public and/or government consideration, whether through the media or by some other means. For CARE, this might not trace back specifically to evaluation findings, because the organization is arranged in such a way that

those involved with public and media relations are not always direct recipients of evaluation reports, nor do they typically participate in the process.

Policy-oriented learning – this refers to attitude change about policy objectives that come about because of evaluation results. In the case of CARE, more than one evaluation report might be needed to effect such a change.

Policy change – related to the above two mechanisms, this form of influence involves an actual shift in policy in the operating environment that results from evaluation findings. Henry and Mark point out that negative evaluation findings might well bring about policy change more than do positive findings.

diffusion – like policy change, this involves a policy shift, but diffusion is about the spread of policy change beyond the operating environment to other contexts. For CARE, this could be other INGOs. This study uses the above-described taxonomy as a checklist that parses the interview data so as to identify which forms of influence evaluations have within CARE, if any.

Coding of Interview Data

To code the interview data, the researcher created a coding scheme based on Henry and Mark's theory of

evaluation influence. She assumed that not all of their categories might appear in this particular set of data. She assigned decision-making criteria to each category so as to discern where to place each piece of interview data. In the instances where there was overlap (for example, the same instance of influence occurring at both an individual level and at an interpersonal level), she coded the piece of data for both categories. She created an "other" category for any data that seemed not to fit Henry and Mark's categories. In the final analysis, this "other" category became the source for speculation about creating additional categories for a theory representing the dynamics of evaluation influence in humanitarian relief INGOs. These coded interviews are then matched with the evaluation reports to identify instances of intended influence. The following is the coding schema for the interview data:

Table 1. Coding Scheme for Interview Data

INFLUENCES		CODE	DECISION-MAKING CRITERIA
Individual	Attitude Change	IND-AC	Any reference at all to individual attitude about a program are coded IND-AC. These are then culled for instances of actual **shift** in attitude resulting from either from participating in an evaluation or from learning of the evaluation's findings.
	Salience	IND-S	Distinct from IND-AC, this goes beyond a change in attitude and is about a priority shift. Example: an issue as an agenda item at a major planning meeting as a result of an evaluation lesson-learned.
	Elaboration	IND-E	Thinking about an issue or situation because of participation in an evaluation or because of having read a report.
	Priming	IND-P	Any instance of an issue rising to the forefront as a result of positioning (such as prominence in an executive summary) or because of the event circumstances themselves (as with the heavily media-covered 2004 tsunami).

Table 1.

	Skill Acquisition	IND-SA	Newly learned or honed skills, such as SWAT analysis in an AAR
	Behavioral Change	IND-BC	Documented instance of change in operating procedure as a result of an evaluation. Should match up with an evaluation report.
Interpersonal	Justification	INT-J	Use of evaluation findings in meetings or elsewhere to make a case for an action or priority
	Persuasion	INT-P	Like INT-J, but goes further: not merely a presentation but an attempt to change others' minds
	Change Agent	INT-CA	Participation in an evaluation, as an evaluator or an interviewee, or as a person reading findings, leading an individual or group to initiate a change effort.
	Social Norms	INT-SN	Example of an evaluation affecting social behavior of those in contact with it.
	Minority-Opinion Influence	INT-MOI	Like INT-P, but specific to a group whose position is that of the minority.

Table 1.

Collective	Agenda Setting	C-AS	An issue emerging from an evaluation that shows up on a meeting agenda for CARE and / or for other INGOs implicated in the effort
	Policy-Oriented Learning	C-POL	One or more evaluation report(s) leads to increased understanding at the policy-making level of the organization.
	Policy Change	C-PC	Like C-POL, but with actual policy change, not merely learning.
	Diffusion	C-D	Policy, program, or practice change resulting from a different evaluation, either at CARE or at related agencies.
OTHER		OTH	

3.5 EXPECTED FINDINGS

Transcripts from five preliminary interviews suggest that more influence occurs at the individual and interpersonal levels than at the collective level. Indeed, the analysis that follows concludes that it is easier to identify examples of influence at the individual level and, to an extent, at the interpersonal level than it is at the collective level. Within these three levels, process-related influence forms are more relevant to some CARE

types of evaluations, whereas evaluation findings, end results, are more pertinent to other forms. The analysis following will put forth a fourth level that clarifies the complex dynamics of influence within a large transnational INGO. It remains to be seen whether the conclusions will be generalizeable to other INGOs working in the same capacity as CARE.

CHAPTER IV

FINDINGS

4.1 Introduction

The findings for this research are presented two parts: findings from the initial review of the 25 evaluations CARE conducted on its emergency response activities from 2000-2005, and the data from the 25 interviews that sounded out Henry and Mark's framework for evaluation influence. The section on the content of the 22 evaluations includes a detailed look at CARE's response to Tropical Storm Jeanne in Haiti. The five pre-interviews resulted in the choice of the Tropical Storm Jeanne evaluation as a highlight; three of the five individuals selected proposed that particular evaluation as a good example of how it "should" be done and of the potential evaluation has as a tool for organizational improvement. Though there is a graphic representation of the interview data included in this chapter, the bulk of the data can be found in the appendices.

The analysis concludes that there are more instances of evaluation influence at the individual and interpersonal levels than at the collective level; that is, the influence of the evaluation reports is not reaching the policymaking

level of the organization or of its peer community. The perception among the interviewees was largely that evaluation report findings are under-used, if they are used at all.

What is perhaps most intriguing about the findings is that the Mark and Henry framework brings out subtle and unintended forms of evaluation influence that not only have implications for how the organization might evaluation more efficiently and effectively in the future, but also point to where in the chain of events the evaluations fall short of provoking policy-level change. The Henry and Mark framework proves to be, for the most part, adaptable to the context of the large transnational INGO, and their "menu," with the addition of a context-specific level of influence, turns out to be a useful tool for identifying instances and forms of evaluation influence. For this study, the framework was awkward as a tool for subsequently mapping "pathways" of influence, but nevertheless was useable for this case. Also, perhaps just as interesting as the instances of influence that emerged were the "non-instances" – that is, the undercurrents hinted at by interviewees that did not fit neatly into the framework's mechanisms. These 'non-events' coupled with the identifiable mechanisms of influence tell an interesting

story of how evaluations affect the organization, as well as how they potentially could.

4.2 Evaluation Reports

The initial meta-evaluation of the evaluation reports served the purpose of assessing the general themes among the recommendations and lessons-learned in the reports, as well as providing a snapshot of the gamut of format and quality characteristics of the reports. The checklist employed for the meta-evaluation, described in detail in chapter 3, drew from four sources relevant to the evaluations' goals and standards. The evaluation reports varied widely in terms of content and format, not surprising given that the evaluators themselves varied a great deal in their experience and style. Also, almost all of the reports were written before CARE International instituted an overarching evaluation policy[2]. Though the variation made comparison challenging, there were recognizable themes throughout the reports. There are three types of evaluations represented:

[2] CARE's Senior Advisor for Design, Monitoring and Evaluation paired with CARE's evaluation specialist to produce a set of evaluation standards for CARE International. The standards are meant to apply to all evaluations undertaken within CARE International. These standards were formally accepted in 2006.

After Action Reviews, also called Lessons-Learned Workshops, are weekend-long reflection sessions taking place right after an emergency intervention. At least a cross-section, if not all, of the individuals involved in the intervention are invited to participate. The aim of the After Action Review is to assess the intervention while it is still fresh in people's minds, think about how to improve on it, and identify opportunities acted upon and opportunities lost. Of the evaluation reports reviewed, 10 were After Action Reviews, with six of those from CARE's response in various countries to the Asian tsunami of 2004.

Real Time Evaluations take place in the middle of an intervention and are intended to be a check-in on how the intervention is progressing and on what might be improved upon. It takes place in the middle of the process and involves just a few individuals (in contrast to an After Action Review, the Real Time Evaluation cannot involve a large number of employees because most are occupied with the intervention. In an emergency situation, the luxury of involving all participants would mean a lesser response). It is meant to inform the rest of the intervention.

Final Evaluations occur after the intervention has taken place and any temporary employees have gone on to other projects. The goal is to capsulate the intervention and to gather information for informing later planning efforts. Final evaluations usually involve an external[3] evaluator. An example of a Final Evaluation:

> **Evaluation of CARE Afghanistan's Emergency Response, September 2002**
>
> This final evaluation of CARE's Afghanistan response resulted in a 64-page final report. An external evaluator and an internal evaluator teamed up to conduct the evaluation. The extensive evaluation consisted of a site visit, telephone and in-person interviews, surveys of CARE staff and of beneficiaries, and a document review. The summary of recommendations makes the substantial length of the final report more manageable for the reader. The evaluation report concludes that CARE's overall response was effective and that CARE is well-received in Afghanistan. Highlights from the recommendations include clarification of lines of authority and roles, especially with temporary staff, and contingency plans for emergency responses.

---

[3] Michael Scriven (1991) distinguishes between internal evaluators, employed by the organization, and external evaluators, coming from outside of the organization. This is an important distinction in any evaluation setting, and CARE is no different. Several interviewees commented that "external" evaluators are often former CARE employees who found that contract work suited them better than working in what is perceived to be a bureaucracy. A few interviewees also remarked that the organizational politics favored external evaluators, but that CARE is a complex organization for an outsider to understand immediately.

A subset of the Final Evaluations, called **Joint Evaluations**, is also represented; in this study, these are uniquely a phenomenon of the Asian Tsunami response of 2004. Several large-scale NGOs, prominent in their response to the tsunami disaster, elected to participate together in a joint evaluation of their efforts. A large goal of their efforts was to determine who did what and to identify competitive advantage, that is, who is better equipped to handle which parts of an emergency for maximum effectiveness. Some of these joint evaluations were products of the ECB (emergency capacity-building) project, a Gates foundation-funded effort to improve the collaboration between the most far-reaching NGOs on emergency issues.

The 22 CARE evaluation reports reviewed for this exercise (all of the evaluations conducted between 2000 and 2005) were, for the most part, lengthy. The reports range from 2-page summaries to 78-page reviews, with the majority of the reports containing 25+ pages. The exception to this is the After Action Review (AAR) summaries, which attempt to capture information from a few days of reflection rather than from the duration of an entire program. Most of the reports contain an executive summary, but in some cases these summaries consist of a few pages highlighting the

report's main lessons-learned, whereas in other cases, the executive summary is a mere paragraph or two generally explaining the study. This suggests that those preparing the reports are doing so without explicit guidelines as to whether or how to prepare the executive summary. The lessons-learned and recommendations themselves vary considerably as far as quantity and depth. Moreover, only a few evaluations make a distinction between "lessons-learned" and "recommendations," and none explain what constitutes a "lesson-learned." A firm understanding of what a "lesson learned" is – and what it isn't – could be a useful construct for CARE. Naming something a "lesson learned" has a different connotation from naming it a "recommendation"; the former implies that the message has been internalized and acted upon, while the latter suggests that the suggestion be followed up with action. As far as methodology, most of the evaluations relied on interviews and document review, with the exception of AARs, which involve interactive dialogue/reflection.

Beneficiaries

The evaluation reports included the occasional call for greater inclusion of beneficiaries in project design: Afghanistan 2002, for example, recommended increasing

consultation with beneficiaries for emergency projects. Sri Lanka/India 2005 noted that the emergency response suffered in instances where beneficiary consultation was sacrificed for efficiency's sake. However, beneficiary-focused recommendations, or accountability to beneficiaries, were the exception rather than the rule in these reports. It should be noted that for the evaluation reports falling outside of the scope of this study (for those projects concerned with long-term development and rehabilitation, beneficiaries are a more central feature in the recommendations). Furthermore, though the evaluations seem at first glance not to focus much on emergency assistance beneficiaries; this could be a function of language confusion. One evaluator's "beneficiary" may be another evaluator's "community," a term with very different connotations. Notably, the Hurricane Jeanne evaluation for Haiti mentions the needs assessment conducted with beneficiaries there as a "good practice" that CARE would do well to replicate elsewhere.

Decision-making

Chain of command, or lines of authority, is one of the most central themes to emerge from the evaluation reports. Moreover, recommendations to specify or clarify the chain

of command cut across regions, type of emergency, and time. Examples of lessons-learned involving chain of command include:

CARE's overall disaster planning has not established clear "emergency" roles, responsibilities and procedures for deploying staff. Deploying emergency staff learned by trial and error. (India 2001)

Clarify lines of authority…national offices should follow up with country offices on reporting deadlines (Afghanistan 2002)
Reporting lines within a country office should be re-articulated as soon as possible after arrival of external emergency response personnel (West Bank/Gaza 2002)

Facilitate clarifications of roles and responsibilities of CO and CERT staff (Iraq 2003)

Divisions of labor need to be clearly defined (DRC 2004)

Lines of responsibility and leadership should be clearly defined (Haiti 2004)

The majority of the evaluation reports reviewed include some version of a recommendation regarding clarifying lines of authority, visits from senior management, and follow-up on responsibilities. The various recommendations related to lines of authority do not all point to the same suggested structure, but it is clear that established lines of communication and reporting are a priority at all levels of emergency response and directly affect the efficacy of the response.

Human Resources and Personnel

Training for staff deployed to an emergency. This involves orientation to CARE and how it is organized, as well as familiarizing deployed personnel with local procedures and operations. Training local staff, using local training methods as appropriate, is also a common theme. The Afghanistan 2002 evaluation report is particularly detailed in training observations and recommendations.

The added workload for disaster response is a recurring concern throughout the reports reviewed:

"Concern has been expressed that many CARE staff now remain with workloads exponentially increased from pre-disaster days…a closer examination of the current division of labour and staff efficiency/motivation would now be useful."
(Haiti Hurricane Jeanne final report)

Also, four evaluation reports bring up the need for having a terms-of-reference (ToR) for every person deployed to an emergency, or having generic ToRs as part of the CO's preparedness plan.

The need for acceptable living conditions for deployed staff was a concern in three reports. Three reports recommended maintaining an active roster of available and

qualified persons for emergencies, suggesting that if such a thing indeed already exists, that is not always the perception in the moment. Finally, three reports recommended having a senior staff person or senior management person visit the site of the emergency as early as possible into the response. This sentiment is echoed in the interview portion of this study; buy-in on the part of senior management is regarded as critical to morale.

Preparedness Planning

In contrast to the interviews, only a few of the evaluation reports mention preparedness among the lessons-learned; one calls for revising the CI emergency manual. The India lessons-learned from the tsunami response stands out as emphasizing the need for better preparedness in several areas, including procurement, policies, and long term strategy. Five reports mention risk reduction and contingency planning as necessities for better efficiency in the future. As with "beneficiaries" above, "preparedness planning" is a term that is just as often called something else, such as contingency planning or risk reduction.

Procurement and Logistics

Procurement received little mention in the evaluation reports, and in general, procurement-related recommendations were specific to the context in which they were observed rather than being generalizeable CARE-wide. The evaluation report for Albania (2000) recommended a permanent procurement capacity. The India earthquake report mentioned the need for a procurement database. Procurement was of greater priority in the interviews than in the evaluation reports reviewed. Many of the logistics observations concerned communications and the need for adequate devices. The Iraq RTE specifically recommended a minimum standard for procurement for critical items such as vehicles.

Finance

One evaluation report recommended a finance manager for the start of any emergency operation. The Iraq report recommends bringing in an external finance manager if resources allow. Some reports included situation-specific suggestions for soliciting funding (e.g. the 2001 Kenya report suggested simultaneously seeking funding for environmental rehabilitation). A couple of reports recommended a CI emergency fund for a more immediate capability in disaster response situations.

Psychological issues

Four evaluation reports mentioned the psychological duress characterizing emergency response staff. The evaluation on Tropical Storm Jeanne, for example, comments that psychological support was an obvious staff need and was late in coming. The Darfur Real Time Evaluation report, similarly, highlighted the importance of counseling both for the displaced Sudanese and for CARE staff for coping with the crisis. The Kosovo After Action Review recommended the continuation of provision of counseling to staff. The Multi-Agency evaluation for Thailand and Indonesia noted the great need for psychological healing on the part of both victims and response staff, noting that current resources are not adequate.

Communications

Communications recommendations range from observations about the need for reliable technology to comments about the criticality of a seamless flow of information. This theme also emerged in the interviews. Though many of the communications recommendations were situation and context-specific, the over-arching theme was that lines of communication need to be established between CARE factions for each and every emergency response. The Multi Agency

Evaluations asserted that coordination among relief agencies is essential to an effective response, remarking in more than one instance that the current level of coordination between agencies is not sufficient.

Security

Not surprisingly, security came up only in the conflict-area emergency evaluation reports, such as Afghanistan and West bank/Gaza. Two reports (Afghanistan, Iraq) recommended that CI develop a security protocol. The Darfur report expressed a desire that security plans be in place prior to a crisis in volatile settings such as Darfur.

Project Plan

Aspects of project planning / preparedness came up frequently in the evaluation reports. Some distinguished between short-term planning and mid-to long-range planning strategies. The need for a situation analysis to precede any response was a recurring theme. The sentiment emerging from the reports is that a recognized emergency preparedness plan would help CARE in responding more rapidly to crises.

Public Relations / Information Management

A number of reports mentioned a desire to cultivate relations with external partners in order to respond more efficiently. One of the After Action Reviews recommended training more individuals to respond to the media's questions, as the few who were trained to respond were fatigued. Very little mention was made of information management, except in the context of the desire to improve internal lines of communication and reporting.

Policy/Advocacy

Policy and advocacy came up frequently in the evaluation reports; they emerged less in the individual interviews. However, one report (Afghanistan) referred to advocacy as an "appropriate" response for CARE due to a perception on the part of other agencies that CARE is strong in that area. The India earthquake evaluation report suggested that advocacy for beneficiaries would provide for informing them of their right to relief and assistance.

Evaluation and learning

Just two of the reports mentioned lessons learned from past evaluations or reviews. Three reports called for building adequate evaluation resources into all emergency plans and budgets; corollary to that, three reports brought

up the need for early and/or consistent data collection as far as emergency response. Finally, having minimum standards or guidelines for monitoring and evaluation was a recurring recommendation.

Effective evaluation report utilization: Tropical Storm Jeanne

More than one interviewee pointed to the evaluation and After Action Review for Tropical Storm Jeanne in Haiti as a model of how evaluation can effectively inform planning and preparedness. What accounts for the perceived strength of this evaluation process?

Timing. The original relief effort required 500 and then 600 staff. Though it seemed important to assess the relief effort early in the response so as to include staff and avoid losing information, the reality was that the staff was stretched to its limits with the response effort. An initial review in January 2005 following the September storm allowed for the participation of a good cross-section of staff, despite the fact that some had already departed. Moreover, the completion of the full evaluation report in March allowed for its use in the Haiti country office' **annual planning event** in April. The report identified resource gaps, such as storage and distribution points for potable water, that the planning session was able to

address for the following fiscal year. The report further provided fodder for **scenario-building** and subsequent contingency planning by painting a visualizeable picture of the situation.

Morale. The Haiti After-Action Review and thorough subsequent evaluation both provided a forum for staff to **reflect** and highlighted **what they had done well** in the response, rather than remaining limited to listing where their response effort had fallen short.

Communicability. In addition to including a local (and francophone) staff person on the evaluation team, CARE Haiti had the final evaluation report translated into French, which allowed for ease in sharing it both with local staff and with partners, such as the UN and other NGOs.

It is interesting to note that neither the interviews nor the evaluations themselves for Tropical Storm Jeanne particularly followed Kirkhart's proposed dimensions for evaluation influence, despite the evaluation's repeatedly having been touted among interviewees as a good example of an evaluation that "worked."

B. Interview data

The interviews, conducted between March 2006 and August 2006, tapped 25 individuals identified through preliminary research and data-gathering as having a range of roles within the CARE international emergency response structure.

Interviewees represented different tiers of the emergency response framework within CARE. Respondents ranged from those on the "front lines" directly involved with the emergency response to those on the executive level. There were 30 formal interviews supplemented with several informal conversations and interview follow-up discussions. Interviews were conducted both by telephone and in person, and averaged about 45 minutes apiece. The interview protocol (see Appendix) provided a loose format for the semi-structured discussions.

The most common refrain from the interviews was the desire for shorter, more pointed evaluation reports. Those on the front lines remarked that they did not have the luxury of time to read **lengthy reports** and do their jobs in the field; those at the executive level commented that they did not need 40 pages worth of information in order to use the reports to make good policy judgments. One person

specifically recommended a maximum length of 20 pages; most wished simply for recommendations to be on top, separate from the main text body. Three interviewees called for the recommendations to be categorized by job responsibilities, such as finance, human resources, security, etc.

Corollary to the length, many interviewees felt that the **distribution** of evaluation reports is inconsistent. Some were not sure whether they receive the reports consistently or not. Very few of the interviewees seemed to know where to look if they wanted to locate a repository of reports; only one interviewee mentioned Livelink (CARE's repository for the reports and other information) specifically.

The **focus** of the evaluation reports was also of concern to many of the interviewees. The overarching sentiment was that they did not have time to read through and pick out the lessons-learned that applied to their specific tasks, nor the time to go through old evaluation reports when dealing with a new emergency.

Furthermore, there was a question among interviewees of **accountability**, of whose ultimate responsibility it is to assure that lessons are, in fact, "learned".

Those interviewees working on a more operational level experience the evaluation reports as being mired in the theoretical rather than having a practical structure for executing the recommendations. They described the reports as often being too "theoretical" or "academic." Simple language was an expressed desire, as was regular translation into French and Spanish.

Several interviewees mentioned a lack of a **learning culture** within CARE, a lack of structure into which learning could be fed and retained. This culture, for many, was about attitudes and behavior rather than about organizational structure. Suggestions included looking to other models perceived as successful, such as that of World Vision International (one person specifically mentioned WVI's comparatively well-organized procurement system for emergencies) or even corporate models.

The interviews yielded very few examples of lessons-learned from evaluation reports that led directly to actions meant to address them. Most interviewees acknowledged that such learning would be great, but had the attitude that it is not realistic. Several interviewees attribute this to capacity; emergency response is 'only a part of their jobs', and they are working beyond capacity

as it is, so they don't have the luxury of sifting through lengthy reports and learning from them.

The desire to be more involved during the evaluation process was a recurring theme in the interview process. Three interviewees remarked that if they had been interviewed for an evaluation, or had been asked to participate in an after action review, they would have been more likely to read the ensuing evaluation report. Another frequent refrain was the issue of accountability; there is no incentive for following up on recommendations, and no penalty for not doing so. Interviewees offered their opinions of who ought to be following up to assure that evaluation recommendations were implemented (it was never the interviewee's responsibility). Interviewees seemed genuinely to want to do their job well and do it better if possible; evaluations were seen as time-consuming and a hindrance, rather than a means to that end.

Henry and Mark Application

Henry and Mark's Levels of Influence framework proposes to provide a sort of menu from which the evaluator can construct a context-specific framework to represent the forms of influence she observes. That is to say that Henry and Mark do not suggest that all of their categories of

influence will be present in all settings. Though all of the categories of influence in their taxonomy were considered in the coding scheme, not all of them were relevant to the line of questioning in the interviews. Coding the interview data to the categories in Henry and Mark's framework proved to reinforce the conclusion that influence is occurring much more at the individual level than at the collective (organizational) level.

The coding scheme for the study followed Henry and Mark's framework item by item. Though the researcher did not expect to find examples of every one of Henry and Mark's categories of influence -- they themselves portray the framework as a 'menu' from which to choose rather than as a definitive model – she included every one of their categories in the initial coding scheme. She used a simple spreadsheet to organize the data. After coding the interview data, the researcher matched the examples of influence with evaluation report recommendations. This offered both a sense of how much time passed before the influence took place, and gave a sense of whether the influence of the evaluation was intentional or not intentional.

Individual Level

The individual level of influence, for Henry and Mark, represents a change in an individual's beliefs or behavior as a result of having participated in an evaluation or read an evaluation report. Inclusion in this category does not preclude an appearance in the interpersonal or collective categories; rather, it is a designation for cases where that is principally the appropriate description. The following examples were of individual-level forms of evaluation influence for the CARE interview data.

Attitude Change

The evaluations of CARE's response in Haiti to Tropical Storm Jeanne yielded an attitude change in CARE personnel who were involved in the response. The country director at the time for Haiti observed this, remarking that when people were in the throes of the emergency response, it was a lot easier to see where CARE's efforts fell short of addressing the overwhelming need. The evaluations, both the After Action Review and the Final Evaluation, shifted people's attitudes, helping them to see that they had done some good work and made a positive difference. The country director's identification of this instance of evaluation influence is difficult to corroborate with the evaluation reports from the After

Action Review and the Final Evaluation, as neither report explicitly states an intent to help workers feel better about their work. But another interviewee, speaking about a different intervention, commented also that it is far easier to make a list of things gone wrong, and that her impression is that evaluation has the power to help people to see what they have done well.

Behavior Change

The evaluation of the CARE India response to the Gujarat earthquake of 2001 yielded traceable behavior change in CARE India's emergency procedures. The recommendations and lessons-learned (the heading for that evaluation report was "major lessons-learned and recommendations") included a call for CARE India and CARE International both to create operating procedures for disaster response and to take steps to beef up disaster preparedness capacity in general. An interviewee from CARE India asserted that the call for better preparedness resulted in trained community task forces for preparedness at the village level, as well as a regular spot for preparedness at annual planning meetings from that point forward. The test for CARE India's improved preparedness came three years later with the tsunami response. The evaluation for the tsunami response found that CARE India

was prepared and had a rapid response with timely staff deployment.

The 2001 earthquake evaluation also criticized procurement for not being effective enough at purchasing and transporting needed materials. An interviewee said that the evaluation process brought the procurement shortcomings to CARE India's attention, and that CARE India had responded to that in two concrete ways. The first was to have vendors identified in several locations, making distance and time less of an issue. The second was to designate the individual in charge of the emergency as the procurement authority. The 2005 evaluation of the tsunami response in India specifically named the pre-identified vendors as a strength in the procurement aspect of the tsunami response. This change occurred at the individual level – one interviewee described how he had changed his practice as a direct result of participating in the earthquake evaluation. The change also occurred at an institutional level, and Henry and Mark's framework does not offer an obvious mechanism to identify this phenomenon. This is partly due to the complexity of sorting out the dynamics and lines of authority in transnational confederations such as CARE. While the individual, interpersonal, and collective levels certainly ring true

for designations that represent CARE, there are country- and regional- office-level relationships that need also to be considered in where evaluation influence occurs.

The 2-day After Action Review response to the 2005 Niger food crisis also yielded behavioral change for CARE. The evaluation report recommended training for accountants. An interviewee from CARE Niger said that that evaluation report helped them to create a standard – they now stipulate that an accountant working with them must be competent in Microsoft Excel.

The Multi-agency evaluation of the response to the 2005 Niger food crisis also led to a behavioral change, according to a CARE Niger interviewee. The report discussed the need for better staff training for emergency preparedness. Now, CARE Niger both prepares staff for the possibility of an emergency and recruits supplementary staff in advance of future emergencies.

Interpersonal Level

No specific instances of the interpersonal level as the primary level of influence emerged from this set of interview data. This is not to say that the interpersonal level is not relevant to the patterns of evaluation influence exhibited at CARE. A few of the interviewees

alluded to interpersonal dynamics; in particular, three interviewees mentioned their perception that individuals sometimes act as change agents, championing an evaluation recommendation when it strikes a chord with them. The implication from the interviews was that this form of influence is episodic and hard to anticipate. None of the interviews yielded a concrete example of this form of influence, so it is not included in the discussion about pathways below. However, as is the case with any such data set, a different sample of interviewees might well have resulted in the emergence of instances of interpersonal influence. It appears to be potentially a relevant level of influence for the humanitarian aid INGO.

Collective Level

The collective level proved challenging for determining where and how influence occurred. CARE is complex in that it is not strictly hierarchical and it has country, regional, and international-level relationships within the organization, as well as a community of peers. The collective level of influence did not seem to fit any of these relationships neatly. For the purposes of the study, the researcher interpreted "collective" to mean CARE-wide, meaning CARE International and relationships

between the 12 CARE International members. The researcher creates an intra-organizational category to capture the layer of influence occurring more broadly than at the individual and interpersonal level but not as extensively as at the collective level.

Agenda Setting

In the interview data, there was one example of agenda -setting brought about by an evaluation. In Haiti, where both an After Action Review and a Final Evaluation took place, the country director affirmed that she incorporated lessons-learned from those evaluation reports (finalized in March 2005) into the agenda for a June annual preparedness planning session. She cited a streamlined potable-water distribution plan as an action resulting from that evaluation-influenced agenda.

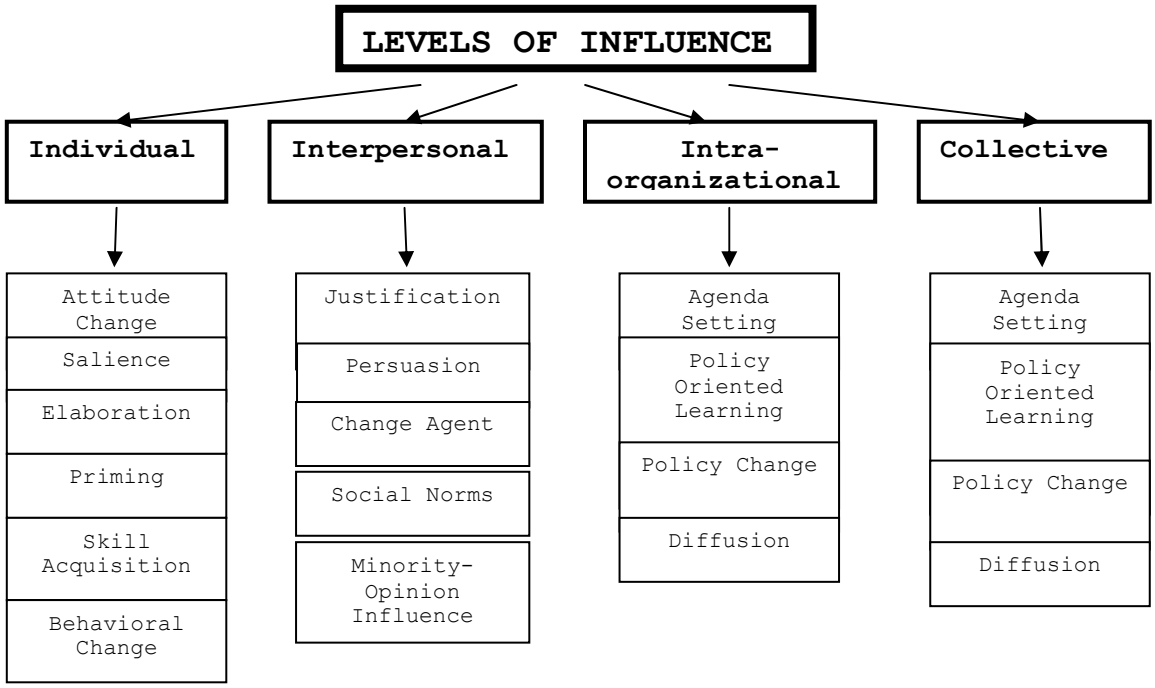
Three other interviewees had the impression that evaluation findings feed into annual and strategic sessions, but could not offer a specific example. Two of those three voiced the opinion that evaluation recommendations become agenda items when they happen to resonate with an individual involved in the process; that is, evaluation results are by no means systematically included in planning agendas.

The above examples show that when we view the CARE case through the lens of the Henry and Mark framework, the few instances of evaluation influence occur mostly at the individual level, with the exception of some agenda-setting on a broader level. The interpersonal level did not surface from the interview data. This could be in part due to the limitations of the interview protocol; the questions highlighted individual and CARE-wide experiences more than it did the interpersonal. Also, the category itself is nebulous when considering the complex CARE organization. Though interpersonal dynamics certainly exist at the level of any given individual office, the category gets murky when considering relating with colleagues overseas, colleagues communicating between developing and industrialized nations, and colleagues between main member organizations connected to the secretariat.

Henry and Mark's framework turns out to be a good tool for the sorting of interview data and evaluation report data. Each of the instances of influence found a home in their sub-categories, notably in the attitude change, behavior change, and agenda-setting categories. The levels of influence also fit the data. However, a CARE-specific theory of evaluation influence would include another level of influence not represented in the Henry and Mark

framework. Large INGOs like CARE often have regional and country-level decision-making centers in addition to the main governing body. In the case of a confederation like CARE, the nexus of the decision-making power does not reside solely with the CARE International secretariat. The regional management units have much of the coordinating role in the event of an emergency response. For this reason, the collective level proved a difficult designation to assign to the interview and evaluation report data for CARE. There were different levels of collective influence that did not fit neatly into the framework. A CARE-specific theory of influence includes a fourth inter-organizational level of influence with the same sub-categories as the collective level, which applies to CARE-wide or inter-INGO examples of evaluation influence. The inter-organizational level is for instances of influence primarily occurring at the country office or regional office level. It is distinctly a different level from a CARE-wide level of agenda-setting or policy change.

Figure 2: Four Levels of Influence for Large INGOs



The Henry and Mark framework includes policy-oriented learning, policy change, and diffusion as potential mechanisms for collective-level influence. None of these three mechanisms emerged from the interview data as characteristic of evaluation influence at CARE. This could be in part because the mechanisms are more suited to looking at policies than to looking at programs. Certainly, the limited time frame for the study (2000-2005) may have precluded the observance of policy change and diffusion, which take time to manifest. Finally, it is

likely that the evaluation findings themselves fall short of reaching policy-level decision-makers.

The intra-organizational level

The researcher proposes an intra-organizational level for a better understanding of the influence dynamics in a complex transnational organization. The intra-organizational level of influence mirrors the collective level, because it primarily concerned with policy-level learning and change, but learning and policies are country- or regional- office wide rather than occurring throughout and beyond CARE as a whole. It is distinct from the interpersonal level because the country and regional offices have their own sets of policies and procedures that are sometimes necessarily region-specific. Different regions have different risk situations; natural resource-poor Haiti, for instance, is chronically plagued with weather and water problems, whereas parts of Africa struggle more with conflict-related emergencies than with natural disasters. Other areas experience emergencies without being constantly disaster-prone.

The Agenda Setting category for the Intra-organizational level involves inclusion of evaluation findings in preparedness or annual planning events. The

designation applies also to brainstorming or knowledge-sharing events within CARE, such as the annual Crosswalk meeting. It differs from the collective level of agenda-setting, which occurs with the executive team level or among the peer group of NGOs of which CARE is a part.

The Policy Oriented Learning category for the Intra-organizational level is about absorption of lessons-learned at the country or regional office level. This manifests itself, for example, in future terms of reference for emergency evaluations that reflect lessons-learned emerging from past emergency evaluations.

The Policy Change category for the Intra-organizational level aims at regional- or country- level policy. An example of this is the decision in CARE India following the 2001 earthquake evaluation to delegate procurement decision-making authority to the individual directly handling the emergency (rather than by default to the country director, or someone else). Diffusion at the Intra-organizational level differs from the Collective level in that the idea diffuses from country to country or country to region, rather than from country to CARE as a whole or from CARE to other INGOs.

Pathways of Influence

Henry and Mark maintain that the categories in their framework are not mutually exclusive and, in fact, can be integrally related as links in a causal chain. They present the framework as an instrument for tracing the 'pathways' of influence, for putting together the links in the chain that lead from an evaluation's initial form of influence to an evolved form. Identifying these pathways then helps organizations to understand the nuanced ways in which evaluations influence them on multiple levels, and also helps them to plan for effective future evaluations.

Because CARE's structure is that of a confederation rather than strictly a hierarchy, the pathways of influence resemble a ripple effect of droplets in a pool of water. The few identifiable pathways of influence emerging from this data are not identical, which suggests that evaluations in CARE are not conducted within a cemented institutional culture of learning. Moreover, it is possible that a different set of 25 interviewees would have identified pathways of influence with a slightly different pattern. So the pattern in CARE's pathways of influence is that influence occurs episodically, as a particular issue

resonates with an individual, rather than predictably and systematically.

An example of this is the evaluation of CARE's response to Tropical Storm Jeanne in Haiti, which involved both an After Action Review and a full Final Evaluation.

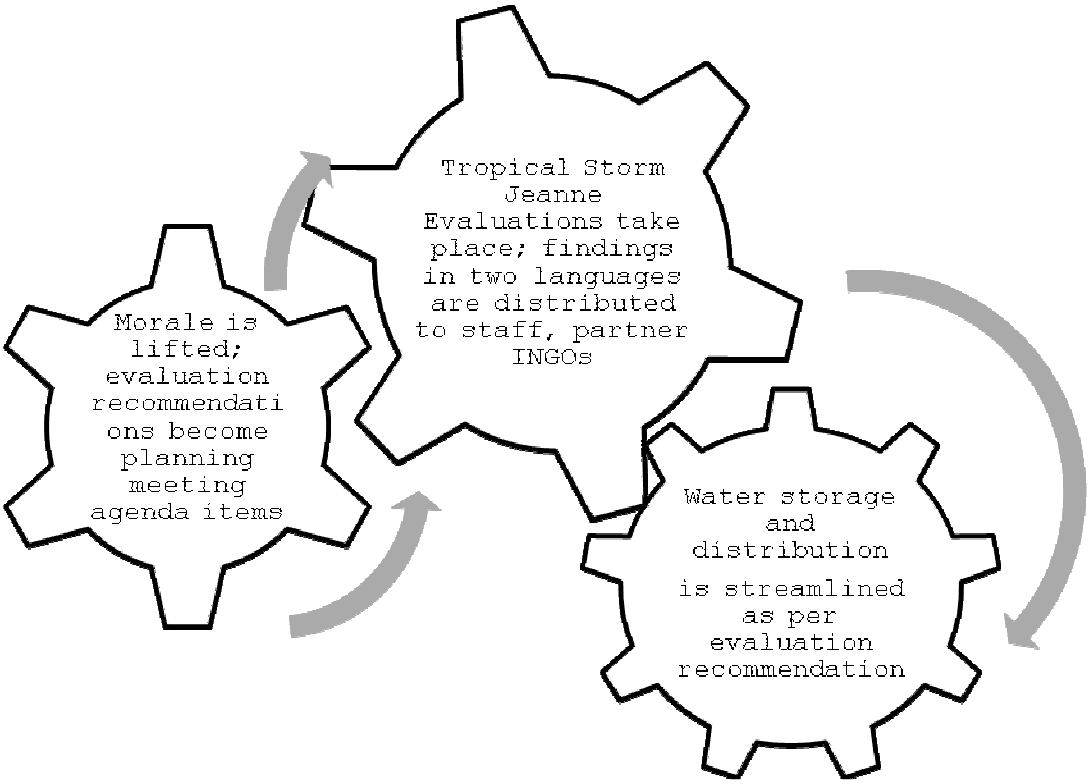
- ↓ The After Action Review yielded a sense among the permanent and deployed staff of having done several things well in the emergency response, whereas in the midst of a crisis that was devastating to CARE Haiti staff and families as well as to Haiti as a whole, it was much easier to see where the intervention fell short.
- ↓ The bolstered morale from the After Action Review then set the stage for the Final Evaluation conducted two months later. CARE staff were willing participants in a process that they might easily have regarded with apprehension.
- ↓ The Final Evaluation, written in English and then translated into French, was distributed beyond CARE to the other INGOs involved locally in the intervention. The staff of CARE Haiti also received the findings from the evaluation.
- ↓ The recommendations from the final report served as material for the preparedness planning effort in June

of that year, as well as a couple of months later for the annual planning event.

↓ One resulting action was the streamlining of the storage and distribution system for potable water for future emergencies in Haiti.

This chain of events shows that the evaluation influenced the organization in both subtle and noticeable ways. An interviewee had sensed an initial reluctance on the part of some staff who were to participate in the After Action Review, a reluctance she attributed to fatigue from months of work in a daunting and depressing situation. The staff may not have participated in the exercise because they anticipated a morale boost, but after having participated, they had a better picture of what they had done well in the emergency response. This example of one evaluation's influence may not have been intentional. CARE Haiti's changing the practice for storing and distributing water, in contrast, followed an intended recommendation of the evaluation report, and is an example of direct and instrumental use of the evaluation.

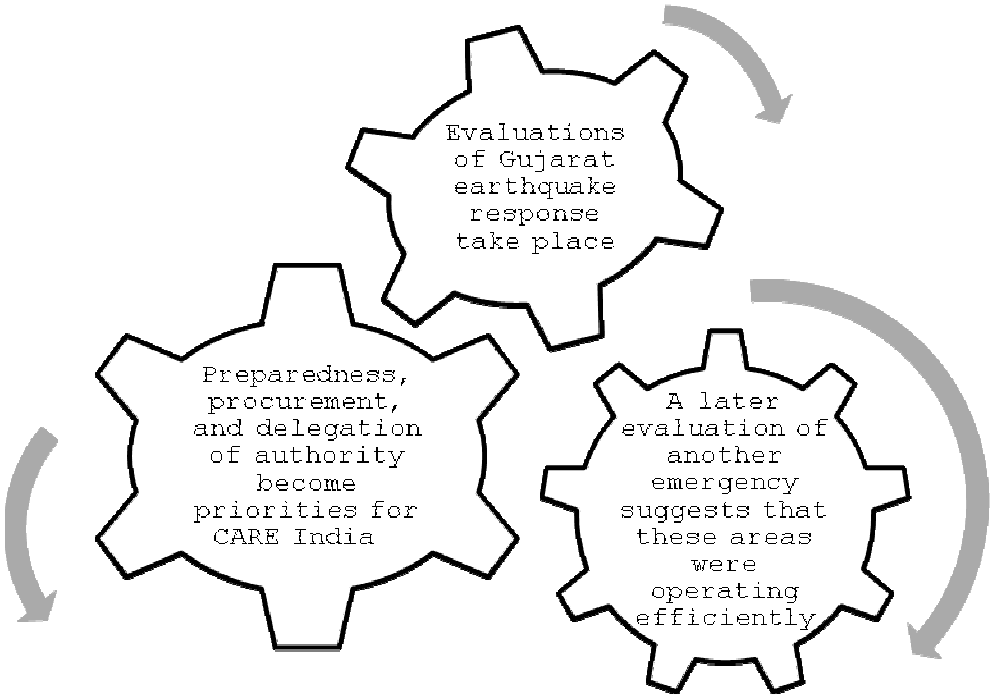
Figure 3: Pathway of influence for Tropical Storm Jeanne



Another chain of events, from the evaluation of CARE's response to the Gujarat earthquake in India in 2001, bears some similarity to the pathway for the Haiti evaluation influence in that there was a similar mix of intended and unintended instances of influence, as well as direct, instrumental forms of influence alongside more subtle examples. The evaluation took place; the process of evaluating and the findings themselves brought preparedness, procurement and delegation of authority to evaluation participants' attention as areas in need of improvement. CARE India took action on all three of these themes, resulting in a more positive review of these

aspects of emergency response in a later evaluation of a different emergency.

Figure 4: Pathways of Influence for India Earthquake Evaluation



Looking at how the pathways of influence shape up instead of merely identifying the types of influence offers insight into how the design, the timing, or the implementation of the evaluation affect how the evaluation influences the organization. These give the organization an idea of where to start in meta-evaluating how their evaluations influence. Identifying these pathways will enable the organization to design ways of verifying the impact of the evaluation.

This study initially tried to use Kirkhart's (2000) Integrated Theory of Evaluation Influence to explain the patterns of influence at CARE. Kirkhart's theory, while it did not codify adequately enough to be useful for such an application, does make a compelling case for the idea of influence versus evaluation use and her matrix of dimensions of influence serve as a launching point for discussion about influence in a particular case, if it is not particularly suited for adapting to individual contexts. Kirkhart, who, like Henry and Mark, espouses the more inclusive term 'influence' rather than 'use' to describe the effect of an evaluation on a policy or program, proposes three 'dimensions' of influence for looking at evaluations. The *intention* dimension considers whether the instance of influence was or was not an intended result of the evaluation process or the evaluation report. Her *source of influence* dimension looks at whether the instance of influence stems from the process of evaluating, or from the results of the evaluation. Her third and final dimension, *time*, offers three periods of timing during which influence might have more or less opportunity: immediate, end-of-cycle, and long-term. All of Kirkhart's dimensions are relevant to the data at hand;

the time dimension is the most difficult to observe in the CARE data.

This observation challenge is in large part due to the short span of time (2000-2005) over which the researcher studied evaluations. (This choice had mostly to do with availability and accessibility of reports prior to 2000). Furthermore, several of the evaluation reports studied were of CARE's response to the 2004 Asian tsunami in several countries, where in some cases there was no precedent of CARE's having operated there prior to the tsunami. So not only had relatively little time passed from 2000-2005, the period of study, but just one country in the study had an emergency after having evaluated a prior emergency. Comparing the evaluations of the Gujarat earthquake response (2001) and the Indian tsunami response (2005) did reveal longer-term effects of the first evaluation that might not have been observable a couple of years prior.

Comparing the two India evaluations over time also highlighted another aspect of evaluation influence: the areas where the evaluation process and findings did not influence the organization as intended. The evaluation report for the Gujarat earthquake response included a recommendation to consult with beneficiaries in designing the emergency response, so as to meet their needs

adequately. The later evaluation also recommended consulting the local community on its needs and wishes, suggesting that that recommendation was not, in fact, a "lesson learned." It is beyond the scope of this study to delve into why some lessons go unlearned and others end up as agenda items for planning purposes, but discovering an instance of a lesson unlearned alongside two lessons learned from the same evaluation shows that the absent examples of evaluation influence can be just as instructive for an organization. The idea of looking at where lessons went unlearned, which could even be the result of negative influence, merits further study.

The following table summarizes the instances of influence identified through the interview data. At first glance, the table looks sparse. This is because many of the interviewees perceived areas of influence that the interviewer could not verify. Only those instances verifiable through multiple interviews or through corroboration with evaluation reports are recorded in the table. Also, there are just three evaluations represented in the table. Many of the evaluations reviewed for the study, most notably those from the 2004 Asian tsunami, do not show up in the table because at the time the study was conducted, so little time had elapsed between the

evaluations and the interviews that it was not possible to verify any perceived examples of influence. The arrows in the table indicate where instances of influence link to become pathways.

TABLE 2. Instances of Influence

LEVELS OF INFLUENCE	Tropical Storm Jeanne	India Earthquake	Niger Food Crisis
INDIVIDUAL			
Attitude Change	X		
Salience			
Elaboration			
Priming			
Skill Acquisition		X	
Behavioral Change	X	X	X
INTERPERSONAL			
Justification			
Persuasion			
Change Agent			
Social Norms			
Minority-Opinion Influence			
INTRAORGANIZATIONAL			
Agenda Setting	X	X	X
Policy Oriented Learning			
Policy Change	X	X	
Diffusion			
COLLECTIVE			
Agenda Setting			
Policy Oriented Learning			
Policy Change			
Diffusion			

CARE's Evaluation Influence 'Story'

The two pathways of influence proposed above show two distinctly different routes from the evaluation to the changes brought about through influence. This is a useful distinction because they are two different cases of different kinds of emergencies, yet they tell a similar story of how influence happens within CARE. Both cases included both an After Action Review, which is by nature a highly participatory exercise involving individual reflection and interpersonal dialogue, and a Final Evaluation. The same is true for a third example, CARE's response to the famine in Niger. For all three examples, Haiti, India, and Niger, the person(s) responsible for getting lessons-learned on the planning agenda for going forward had participated in (or led, in one case) the After Action Review, and were involved at least as interviewees in the Final Evaluation. This may in part explain why the influence reaches the intra-organizational level, but consistently falls short of effecting change at the collective organizational level or of diffusing to peer or partner INGOs: those responsible for setting the agenda at the collective organizational level are removed from the front lines of the process of evaluation and have only lengthy reports from which to sort through possible priorities. The lessons consequently go unlearned.

It surprises no one that this study finds many more examples of lessons unlearned than of lessons learned. There are dozens of possible reasons for this. The lessons learned, however, are examples of intended, direct, instrumental influence flanked by unintended and indirect influences. These examples lend themselves to a sort of theory of influence specific to CARE (and possibly other similar INGOs) in the emergency response context, a theory of influence that can function as a tool for facilitating evaluation influence at the planning stage of the evaluation.

CHAPTER V

DISCUSSION AND CONCLUSION

This study shows that the consequence of evaluation reports goes beyond use to encompass influence, which includes both direct, intended evaluation results and indirect, unintended effects. In the case reviewed for this study, evaluation reports were not merely window-dressing; in addition to upholding accountability, they affected individual attitude and behavior, relationships, and country- and regional- level change. The factors leading to these changes, in large part the format and timing of the evaluations, can inform future evaluations so that they will have greater positive influence.

So What?

The 1990s brought more media attention to INGOs and their work than ever before. Well-publicized scandals in the disaster relief community at the time heightened both the public's concern and that of donors over INGOs' responsible use of donated funds in responding to emergencies. Non-governmental organizations are experiencing a call to action to institute or improve on their existing systems of accountability. This has

resulted in a rise in evaluative activity on the part of humanitarian aid organizations.

But the question remains as to whether this newfound energy and effort toward greater and more rigorous accountability results in better or more responsible work and use of funding on the part of these INGOs. Evaluations cost time and money at the very least, two resources which are by definition scarce for an INGO. Any dollar wasted on an ineffective evaluation is a dollar that might have gone toward emergency relief supplies. So it is important to ask whether evaluations do improve INGOs' practice (and thereby by extension help them to fulfill their mission and make the world a better place).

It is not hard to believe that there were far more cases of lessons *not learned* in the evaluations reviewed in this study than of lessons *learned*, that is to say recommendations that resulted in action and subsequent improvement. At first glance, it seems rational to expect that a chain of actions would look something like this: a) CARE commissions an evaluation of its response to a particular emergency b) a team of evaluators conducts the evaluation c) the evaluation team produces a report that includes recommendations for how to improve on current practice d) personnel diligently read the disseminated

report and do their part to make sure that the recommendations are carried out. Of course it does not work quite like this. If it did, there would be no body of literature spanning three decades on evaluation utilization.

The literature on whether and how evaluation findings are used has evolved to consider a range of ways that evaluation might affect an organization and its operating environment. This sort of consequence, or 'influence,' looks past evaluation use to how evaluation might affect even the broad policy climate. As the CARE case also shows, the evaluation influence does not have to be direct or intentional to have a positive and worthwhile impact on an organization.

As accountability gains prominence as an area of importance for INGOs, the question of whether and how the evaluations that make up the backbone of the accountability system influence the INGO and its operating environment looms large. In emergency response, improving one's practice can have such grand effect as saving a life. It is also important to consider whether evaluations influence the organization in smaller, subtler ways, even if the bulk of the lessons appear to go unlearned. This study has used Henry and Mark's framework for evaluation influence to look

at the case of CARE, a prominent INGO in the area of emergency response.

Limitations of the Study

Most, though not all, of the interviews conducted for this study were with employees of CARE USA. Though this is in part a function of CARE USA's comparatively significant size and role in emergency response, it is important to acknowledge here that the research findings represent CARE USA more accurately or more thoroughly than they do some of the smaller CARE members. Also, although all of the evaluations conducted between 2000 and 2005 were reviewed for this study, it is important to consider that a disproportionate number of them were from CARE's response to the Asian tsunami of 2004.

As with any qualitative research, there is a level of subjectivity to this study, particularly in its heavy reliance on interviews that capture individuals' perceptions. A single person conducted the interviews and culled the evaluation reports, leaving room for the possibility of a one-sided interpretation of the interview and report data. Also, selection of interviewees relied on 6 people's informed opinions about who would comprise a representation of those at different levels of the evaluation process. The researcher conducted five pre-

interview interviews as a way of refining which questions to ask and ascertaining whom to interview. CARE's Coordinator for Quality, Accountability, and Standards then helped the researcher to make the final selection of whom to interview. Though the selection of interviewees was careful and thoughtful, 25 other interviewees might have painted a different picture of how evaluation works at CARE.

Discussion on the Findings

Henry and Mark's framework for tracing evaluation influence showed itself to be a useful tool for developing a context-specific taxonomy that represents how and on what levels evaluations affect people, places and policies. In the case of CARE, "context-specific" meant adding an Intra-organizational level of influence. This helped to distinguish between examples of influence that are collective at a country or regional level from examples that are collective at an organization-wide or peer group-wide level. The subcategories for the Intra-organizational level mirror those for the Collective level.

Using this modified framework to sort the interview responses and compare them to the evaluation reports yielded a picture of how evaluation influence comes about in CARE's emergency response unit. There were three

examples of direct, instrumental evaluation use at the country and regional level:

- The evaluations of CARE Haiti's response to Tropical Storm Jeanne yielded an eventual streamlining of CARE Haiti's potable water storage and distribution system.

- The evaluations of CARE India's response to the Gujarat earthquake resulted in the delegating of authority for procurement decisions for future emergencies, as well as an established list of identified local vendors of disaster relief materials.

- The evaluations of CARE India's response to the Gujarat earthquake prompted CARE India to develop a concrete disaster preparedness plan.

There were also examples of more modest, not-necessarily-intended use:

- The After Action Review for CARE Haiti and Tropical Storm Jeanne reassured emergency staff that they were doing good work and making a difference. This set the stage for the final evaluation two months later.

- The Tropical Storm Jeanne evaluations became the basis for the agenda of a preparedness-planning meeting three months later, and an annual planning meeting two months after that.

It is easy to point to numerous recommendations from the evaluations in the study as lessons un-learned, or recommendations for which there is no evidence of resulting change. Of course, some of these are cases of countries where there has not yet been another emergency to show where the emergency response has improved due to recommendations from evaluations of prior emergencies. Some

are also cases where a different sample of interviewees might have had different experiences of evaluation, resulting in a different set of examples of lessons learned and unlearned.

There is one emergency setting for which two evaluations, one from an earlier (2001) emergency and one from a later (2004) emergency, show that some lessons were learned whereas others were not. Paralleling the Gujarat earthquake evaluation findings with the tsunami India evaluation findings suggests that though there were a couple of areas (mentioned above)where the 2004 response effort improved from the 2001 response, there was an area where the lesson remained unlearned: both evaluations concluded that it is essential to consult with the beneficiary community about the response plan before going ahead with the emergency response, and both recommended that this be implemented in the future.

Other examples of evaluation influence not occurring or not occurring systematically emerged from the interview data. Two interviewees who are country directors said that evaluation findings are sometimes considered in annual planning events, but not systematically. Both speculated that some evaluation recommendations find their way onto annual planning agendas if they happen to resonate with an

individual who champions them. A member of CARE's Impact Measurement and Learning team echoed this opinion.

Another example of missed opportunity for evaluation influence lies in personnel participation in evaluations. Three interviewees who had task-related jobs (rather than country director or other managerial positions) expressed the desire to be asked to participate in After Action Reviews or asked to give interviews for Final Evaluations. The interviewees lamented the lost opportunity to reflect or to offer an opinion about the emergency response.

Pathways of Influence

Mapping out the instances of influence into patterns of chains of influence shows in the case of CARE that evaluations have greater and farther reaching influence than might seem on the surface to be the case. When asked directly about whether "lessons-learned" are actually learned, most of the interviewees admitted to seldom if ever reading the evaluation reports and expressed skepticism concerning whether lessons actually get learned. One person went so far as to say that in an emergency, planning happens immediately and speedily, and no one is going to reach back to the recommendations from a previous evaluation for insight in such a moment of urgency. The pathways of influence illuminated in this CARE case show

that the evaluations may have a greater effect, sometimes in subtle ways, than the interviewees think that they do.

Both of the pathways of influence traceable in this CARE case show the process part of the evaluation to be important to its overall influence. Involvement in the process of evaluating created "buy-in" and made key people sit up and take notice of necessary and doable changes that could be made. The After Action Review offered a place to reflect and gave deployed personnel a forum for discussing their experiences. In the cases of Haiti/Tropical Storm Jeanne and India/tsunami response, the After Action Review set the stage for successful Final Evaluation activities by creating good will on the part of personnel for participating in the evaluation activity. Thus, though a substantial number of evaluation recommendations do not become action items, the evaluation process, particularly in the case of After Action Reviews followed by Final Evaluations, have much value in that they can boost participants' morale in the midst of a depressing disaster scene. Participating in an evaluation can motivate a person to pay attention to the evaluation findings.

Lessons (Un)Learned

Identifying the pathways of influence sheds some light on why some evaluation lessons-learned receive attention

and others do not. As an illustration, the case of the Haiti Tropical Storm Jeanne evaluations culminating in a change in procedure for handling potable water was an example of a recommendation for better preparedness that was easy to execute. Furthermore, the preparedness planning meeting coincided with the availability and dissemination of the evaluation recommendations. The ascertaining of procurement vendors in India followed much the same pattern. Lessons are learned when the timing of the evaluations and evaluation reports is good (coincides with planning meetings without interfering with the emergency response itself), when the evaluators are well-received or well-regarded, or when one particular lesson resonates with an individual who makes an extra effort to see the recommendation carried through.

Lessons that go unlearned, however, do so for myriad reasons. Five of the interviewees mentioned, for example, the absence of a 'culture of learning' or 'culture of accountability' at CARE as a culprit for not capturing lessons-learned, and two more interviewees called it a lack of 'institutional memory.' It is probably unfair to say that CARE lacks entirely a culture of learning. After all, CARE USA has a team of four persons whose function is to strengthen CARE's accountability to itself and others

through inquiry and learning. It may be that evaluations are not fully realized as tools within the learning structure. But it would appear that evaluation of emergency response is very slowly integrating into the culture at CARE as a means of reflecting on past practice and improving on future practice:

- Part of the decision to look at the 2000-2005 evaluations for this study rather than start farther back than 2000 is that it is only in the year 2000 that there are consistent records and evaluation reports for each emergency response in CARE's repository.
- CARE International adopted a carefully conceived set of evaluation standards in 2005 that inform the design and execution of all evaluations of policies or programs.
- By 2005, After Action Reviews accompany all of CARE's evaluations of emergency response activities, suggesting that giving deployed personnel a forum for reflection and helping them to identify the positives in the impact they have had is a shifting priority for CARE.
- The 2004 Asian tsunami and the 2005 Niger food crisis were events for which CARE participated in a multi-

agency evaluation, suggesting a comfort level in sharing evaluation findings with peers.

An emergency can be a grisly, depressing scene. When the needs of the afflicted are as basic as water and shelter, a person's instinct might be to put a response plan in place as rapidly as possible so as to begin to mitigate the suffering. To paraphrase one interviewee's sentiments, "you're not going to go poring back through evaluations of former emergencies for lessons-learned before at a time like that." Another interviewee pointed out that some evaluation recommendations are easier to carry out than others are in an emergency situation. The interviewee suggested that sometimes it is a matter of capacity – it is not that CARE is ignoring a recommendation, but that it does not have the capacity to respond to it fully. He pointed out that for the majority of staff deployed in an emergency, emergency response is one sliver of their job – most are not devoted to emergencies full-time. This is the excuse many interviewees gave for not reading lengthy evaluation reports; emergencies are but one small portion of their jobs.

CARE's Evaluation Influence vis-à-vis Its Peers

Though it is not a main aim of this research to compare evaluation influence between two or more INGOs in depth, it is important to consider CARE in the setting of its INGO peers. This is essential because emergency response frequently is a team effort drawing simultaneously from the strengths of each INGO. The effectiveness of the emergency response is greater than the sum of the parts – the individual INGOs – that participated. This is clear from the Multi-Agency Evaluations reviewed for this study; the agencies participating in the evaluations each had strengths that contributed to the response in ways that the other INGOs could not on their own. Determining how to increase the influence of evaluations in CARE not only has implications for similarly large, decentralized transnational organizations, it means more useful evaluation reports for CARE's peers and partners in emergency response. The Asian tsunami of 2004 brought with it a great deal of collaboration among INGOs because the damage wrought was so severe as to require INGO assistance on a massive scale. Some of the humanitarian aid agencies collaborated on jointly evaluating their response efforts for the tsunami and for the 2005 food crisis in Niger. It is difficult to compare the respective findings for each agency, as each had a distinct role based on its strengths

and resources in the emergency response. But it is significant of a shift in culture that the agencies found value not only in evaluating their respective activities, but in evaluating the inter-agency collaboration with the hope of improving on it. Time will tell how these evaluations influence that collaboration.

Though the other agencies[4] involved in the Multi-Agency Evaluations have a confederation structure like CARE (most have a more traditional hierarchical structure), the framework developed in this study and the ensuing theory for how pathways of influence work in the CARE setting are likely to be applicable for these other major players in the emergency response arena because they face similar challenges of different levels of management, from the front lines to the country and regional level to the executive and policy level. All of these agencies work in development as well as in humanitarian relief, and all of them have been prominent figures in recent emergency situations. How they compare in terms of evaluation influence within and beyond the agency is beyond the scope of this study. However, an Overseas Development Institute

[4] World Vision International, Oxfam Great Britain. Catholic Relief Services, Save the Children, International Rescue Committee, and Mercy Corps

(ODI) report (2005)[5] found that evaluation is a particular area of strength for CARE, though there is not a systematic way to integrate lessons-learned into policy formation for the organization.

CARE-Specific Recommendations

A couple of interviewees expressed a wish that they could participate in evaluations from other countries and regions within CARE's system so as to learn from their successes and failures. Breaking down barriers in encouraging a dialogue of ideas and experience will promote a learning culture (Garvin 2000). This sort of exchange can be instituted into the regular evaluation cycle.

The interviews highlight a particular strength within CARE: employees really do care about doing their jobs well and are willing to work hard at that in the most adverse of circumstances. When lessons go unlearned, it is not so much a problem of indifference as it is of attitudes, time

[5] The report, commissioned by the CARE International Secretariat, addressed CARE's perception that its emergency response needed to be assessed and improved upon. The report compared CARE to World Vision, Oxfam, Médecins Sans Frontières (MSF), the International Rescue Committee (IRC), Save the Children, and the International Red Cross and Red Crescent Society across four benchmarks: timeless of emergency response, appropriateness of response, proportionality (in scale) of the response, and effectiveness in achieving objectives. The review found that CARE has not grown as its peers have grown in terms of its capacity to respond to emergencies.

availability, and confusion over whose responsibility it is to assure that change occurs. The following are recommendations for building a stronger culture for lesson-learning, where evaluations could be engineered to have greater influence:

1. Standardized Template for evaluations. The evaluation reports reviewed are not uniform in terms of form, content and methodology, although the more recent evaluations are beginning to resemble each other. A more standardized format would be helpful for those looking to skim the report rapidly. It would help evaluators in ensuring that their outputs were in line with CARE's expectations. Fields on the format could be linked to a searchable database to allow easy access to lessons learned in a concise format, either from individual evaluations or in the form of a synthesis (e.g. a summary of recommendations relating to human resources over the past two years).

How: The recommendation is for a **standardized Terms of Reference and evaluation format** that would include:

- qualifications of the evaluator including whether he/she has ever worked for CARE and/or "knows" CARE,
- methods used to conduct the evaluation,
- minimum baseline data collection,
- evaluation findings for a list of subcategories,

- o findings and recommendations. This last category would have a clear distinction made between a "lesson learned" and a "recommendation," with a section for "good practices," or positive lessons-learned that ought to be replicated.

2. Template or guideline for AARs: The After Action Reviews were perceived positively by most of the interviewees as a means of identifying lessons-learned through evaluative reflection. A thorough how-to for conducting one, or at least reporting on one, would facilitate the use of AAR findings.

How: The AAR should take place early enough that those who responded to the emergency are still there, but late enough that the AAR does not interfere with the response effort. The review should consist of reflection both on the process of the response and on the end result. In addition to the individuals directly involved in the response, representatives from human resources, procurement, logistics, security, and external relations should be invited to participate. Each attendee should receive at least a summary of the AAR notes and recommendations. The

facilitator should be competent in the relevant language for the AAR.

3. Yearly synthesis of priority themes to coincide with planning cycle of CARE: It is clear from the interviews conducted for this study that CARE employees, like those of most nonprofit organizations, are time-starved from the operational level all the way up to senior management. Moreover, the lengthiness of the evaluation reports dissuades people from reading them and from wading through text to identify recommendations relevant to the individual's job. An annual synthesis of important themes and identification of themes on which to focus for the year would assist in shaping CARE's policy and planning agenda. Several of the individuals interviewed envisioned this yearly synthesis as coinciding with the end of the calendar year in December, in anticipation of January planning sessions for the following fiscal year. Others saw the Annual Operating Plan meetings as the forum for such a synthesis. This synthesis, with follow-up from previous syntheses, would be appropriate at bi-annual ERWG meetings, as well.

How: the person responsible for quality assurance within CARE's emergency group would lead a synthesis exercise in November and December of each year. (It is important that the person leading the exercise be thoroughly familiar with the responses reviewed, and that the person be high enough within CARE to assure buy-in from all concerned. Significantly, the main researcher for this study triangulated the scan of evaluations by having another researcher, new to CARE, categorize the data as well. The two scans did not match up in terms of categorizing findings; presumably, identifying priority areas would be still more difficult for an individual not thoroughly versed in CARE's emergency response programs). The exercise would involve reviewing any evaluation reports from the year leading up to that point and prioritizing the recommendations listed in each, identifying who should be responsible for follow-up for each of the recommendations. The synthesis would be prepared for planning meetings in January and would not exceed 10 pages.

4. Cover sheet for evaluation reports that can feed into a searchable database: Individuals perceive evaluation reports as too cumbersome to be practical for incorporating specific lessons-learned. The reports are lengthy, and

recommendations targeting a specific area, such as human resources, get lost among all the other recommendations. A "cover sheet" for evaluation reports, to be completed eventually by the evaluator preparing the initial report, would categorize lessons-learned into areas of specialty, such as human resources, external relations, procurement, etc, so as to facilitate the use of the report findings by individuals who are responsible only for a slice of the findings. Though the evaluation reports are now easily accessed in their Livelink location, it is more of a repository than a database. A **database** would allow searching by region, or disaster type, or by job sector, specifically: human resources/personnel, finance, procurement/logistics, advocacy, security, and monitoring/evaluation.

How: Potentially, such a cover sheet could eventually be incorporated into a searchable database allowing users to search for evaluation reports containing information relevant to their jobs. The cover sheet would be no more than two pages long and the evaluator submitting the report would complete the cover sheet. Topics included on the sheet:

- Short abstract of the emergency context
- Time frame of the response and evaluation(s)
- Country & Region

- Sector(s)
- Type of evaluation (AAR, RTE, joint evaluation, etc.)
- Type of emergency (conflict, natural disaster, slow-onset, etc.)
- Lessons-learned categorized by job function (logistics, procurement, human resources, external relations, etc.)

5. Policy on internal vs. external evaluators: The interviewees, when queried about the plusses and minuses of using internal versus external evaluators (internal to CARE vs. external to CARE), responded predictably that while external evaluators sometimes have too large a learning curve in terms of understanding how CARE works, they bring a fresh perspective. Internal evaluators, on the other hand, know how CARE is structured but can be in a politically awkward situation within the organization or can lack perspective. They may not always ask the tough questions. Most interviewees agreed that a team of evaluators, internal and external, is ideal when possible. Creating a "bank" of external evaluators who are familiar with CARE and who are known to be competent would facilitate this. This already exists, in a way; human resources maintains a roster. The recommendation is that this resource be formalized.

How: Though "prior CARE experience" is a criterion for hiring evaluators, there is a perception among interviewees

that there is great disparity between external evaluators in terms of their understanding of CARE and their experience as evaluators. Formalizing, even training, a group of emergency evaluators could assure that they know CARE, know the desired format and content of the evaluation, and know how to complete the "cover sheet" (see above).

6. Learning opportunities: Several of the interviewees were of the opinion that other countries and regions could learn from their emergency response experiences, and vice versa. Inviting staff from other countries and/or regions to After Action Reviews, planning meetings, participating in evaluations as team members and other such events on a rotating basis might enable valuable sharing and reflection.

How: Budget for at least one individual from a neighboring region's CARE office to sit in on each After Action Review. Share that individual's reflections and reactions widely. Also, systematically translate evaluation reports into French and Spanish.

7. Clarify language for "lessons-learned," "findings," "recommendations," "best practices," and the like. In some evaluation reports, findings and lessons-learned are two

different things; in others, recommendations differ (follow) from lessons-learned. "Best practices" and 'lessons learned" are not so different, Patton reminds us (2001). This is not an issue unique to CARE, certainly[6], but certainly greater specificity of language would help to foster a better learning culture. There were numerous comments in the interviews indicating confusion about who is responsible to see that evaluation recommendations are carried out. This suggests that "lessons-learned" is a misnomer, and that "recommendations" is more accurate. Furthermore, perhaps a format specifying how to execute the recommendation and who is logically responsible for seeing it through would make it easier to trace.

Further Study

This study was limited to one division of one INGO, and relied on personal recollection of evaluations that had already taken place. Certainly, comparing two or more NGOs would be a rich way of identifying where commonalities in trends of influence might lie. It would also be interesting within one organization to study an evaluation from its inception through a period of time following the

[6] World Vision International compiled an internal document, a master list of the year's "lessons-learned" in 2006. The list had two columns: the first was the lesson learned, and the second was *a recommendation* for how to act on the lesson learned.

final evaluation, or even through another emergency. The study was limited in that the available data represented only one example where there were two separate emergencies within the time period of study, allowing for observation of which recommendations resurfaced and which seemed to have been addressed. It would be interesting and valuable, also, to take Henry and Mark's framework as it applies to humanitarian aid agencies, and to use it to plan an evaluation with the framework as a basis with the intent of maximizing the influence that the evaluation will have, and then trace whether it is in fact possible to plan an evaluation to have greater influence.

Conclusion

Evaluations of humanitarian aid missions have the potential to affirm and bolster staff morale, discover and increase good practices, and highlight areas for improvement. CARE has experience with effective evaluation utilization, and with some changes to its structures and systems, has the capacity to encourage a culture of learning while putting the framework in place for improved practice. Annual prioritization and synthesis of lessons-learned, a searchable database, standardized evaluations and reviews, and increased learning opportunities are simple steps that can lead to a host of practical

improvements and attitude shifts, as indeed evaluations already have in isolated instances. Expanding the definition of evaluation use to influence, allowing for unintended and indirect forms of influence, reveals important ways in which evaluations add value to the humanitarian aid agency's practice. Understanding these patterns of influence better, by identifying how they link and form pathways, shows in this case that evaluations can, with planning, have even greater influence and bring out the latent potential of the agency's learning cycle.

It is sure that NGOs are being held to ever-greater standards of accountability, and organizations are even popping up to help them to do so. It is less apparent whether these tightening standards are resulting in evaluations that help a policy or program do its job better; that is, make the world a better place. Development of a theory of evaluation influence for humanitarian assistance context takes a step toward advancing theories of use and influence into a practical means of planning for and implementing evaluations. In the world of humanitarian relief, this could mean better practice and saved lives.

APPENDIX A

LIST OF EVALUATIONS REVIEWED

TITLE	DATE	TYPE	NOTES
1. Evaluation of CARE Afghanistan's Emergency Response	Sept 2002	Final Evaluation	One external, one internal evaluator; site visit, document analysis, interviews of CARE staff and beneficiaries; 64 pages
2. Independent Evaluation of CARE's Humanitarian Response to flooding Resulting from Tropical Storm Jeanne in Haiti	March 2005	Final Evaluation	Two external evaluators and one internal evaluator; document review, field visits and observations, semi-structured interviews and focus groups; 28 pages
3. Joint Independent evaluation of the Humanitarian Re3sponse of CARE, CRS, Save the Children and World Vision to the 2005 Food Crisis in the Republic of Niger	Nov 2005	Final Evaluation / MAE	Six evaluators, two internal; document review, observation and semi-structured interviews and focus groups; 40 pages
4. Final Report, CARE International's Humanitarian Response to the Darfur Crisis	June 2004	RTE	Internal evaluator; interviews and document review; 13 pages

5. Executive Summary and List of Recommendations, CARE International's Humanitarian Response to the Iraq Conflict	Sept 2003	RTE	Three internal evaluators;
6. CARE International Real-Time Evaluation of the West Bank-Gaza Crisis	May 2002	RTE	Internal evaluator; 10 pages
7. Food emergency, Southern Africa: Lessons Learned Workshop Report	June 2003	AAR	Group sharing sessions, SWOT analysis, intro to SPHERE as a monitoring and evaluation tool; 10 pages
8. Sri Lanka and India Multi-agency Evaluation	July 2005	Final Evaluation/ MAE	Team of 5 evaluators, including an external evaluator as team leader; 52 pages
9. Multi-Agency Evaluation of Tsunami Response: Thailand and Indonesia	July 2005	Final Evaluation/ MAE	Team of 5 evaluators, including an external evaluator as team leader; document review, focus groups and interviews; 62 pages
10 CARE West Bank and Gaza emergency Programme Strategic Review	May 2004	Final Evaluation	Team of 6 evaluators; interviews, focus groups, observation; 78 pages

11	Lessons Learned Workshop, CARE Ethiopia	Oct 2003	AAR	Group sessions based on a 2000 lessons-learned report; one-day workshop
12	CARE Ethiopia Lessons Learned from Year 2000 Emergency Operation	Dec 2001	Final Report	33 ages
13	Hurricane Jeanne: CARE Haiti's Response to the emergency lessons Learned Workshop / After Action Review	Jan 2005	AAR	Internal facilitator; chronology of events, small working groups; lessons-learned divided into positive and negative; 24 pages
14	India Earthquake Executive Summary	March 2001	Final Report	43 pages
15	Kosovo Crisis Lessons Learned Review		AAR	61 pages
16	India Tsunami Response After Action Review	April 2005	AAR	2 days of reflections with selected staff; desk review, gender audit and quantitative survey; 24 pages (60+ pages with annexes)

17	CARE International in Indonesia, Aceh Tsunami Response After Action Review Workshop	April 2005	AAR	
18	Tsunami After Action review, Garowe	March 2005	AAR	
19	Tsunami After Action Review, Colombo	2005	AAR	3 days, lessons learned session plus concepts and approaches sessions – 44 attendees
20	Tsunami After Action Review, Thailand	2005	AAR	
21	Lessons Learned Workshop, Niamey – Executive summary	2005	AAR	
22	Main Lessons Learned, CEG/CARE Sudan Engagement	Dec 2005	Final Evaluation	External evaluator ; this "main lessons" document 10 pages long

APPENDIX B
MATRIX OF LESSONS-LEARNED AND RECOMMENDATIONS

AREA OF FOCUS	Lesson Learned or Recommendation
Beneficiaries	1. A CARE Haiti gender policy with strategies and guidelines would be useful. –Haiti 2005
	2. Develop documentation to inform communities of changes in intervention, such as food distribution. 3. CARE staff identified needs rapidly with communities from the start and worked with coordination community to organize a response to those needs. This should be replicated.–Haiti AAR 2005
	4. Train community youths in repair and maintenance of hand pumps and provide repair kits to each trained youth. 5. Psychosocial programming for adults should be very closely tied to viable livelihoods and housing programs – group counseling sessions and linking therapy to other community interventions should be emphasized. 6. Both CARE and World Vision should consider including alcoholism prevention and counseling components into their psycho-social programming. 7. Be sensitive to the needs and views of the affected community when building shelters in India. 8. Urgent -- undertake repairs to all the soak pits in India. 9. Undertake research and study to develop guidelines for ensuring community-driven response. –India and Sri Lanka MAE 2005

	10. Urgent – watsan improvement in settlements. –Thailand and Indonesia MAE 2005
	11. Beneficiary selection can be improved through a long-range approach rather than a short-term or interim approach. Also – coordination with other agencies on this is desirable. –WBG Strat Review 2004
	12. Communities were not directly involved in the planning of emergency responses. 13. Government obstacles to beneficiary selection –Ethiopia LL 2000
	14. Improve targeting mechanisms to account for environment with high level of uncertainty and unreliability in terms of relief. 15. Build in community contribution and participation whenever possible –South Africa LL 2003
	16. Involve informal women leaders and indigenous people with community and local knowledge in the planning, distribution, forwarding, receiving, and benefit of project inputs. -India Tsunami AAR
	17. Ensure that communities participate meaningfully in relief from day 1. –Indonesia/Aceh AAR
	18. Needs assessments did not always capture the communities' changing needs. If necessary, second needs assessments should be developed. 19. There was community fatigue regarding assessments. Coordinate better in the future. –Garowe AAR

	20. Need for better definition of "affected communities" and better identification of beneficiaries. 21. Staff should enter villages with a process of community analysis rather than a prescribed set of actions. –Thailand AAR
Preparedness and Planning	1. Contingency planning should occur and should include a plan for terrorist attacks. –Afghanistan 2002
	2. Recommendation for using development programming to enhance both preparedness and prevention. – Haiti 2005
	3. Warning system is adequate, but action/response late and under-scale. – Niger MAE 2005
	4. Develop risk management plan and adapt EPP procedures and guidelines. 5. When [the emergency] hit, there was an early warning system for drought only. In this future, this should be expanded, based on risk assessment. –Haiti AAR 2005
	6. CARE must find a balance between participatory approaches/relying on local capacity and fast, scaled-up response. 7. CARE needs to build technical competence and credibility in shelter provision in India. –Sri Lanka and India MAE 2005
	8. Develop national organizational disaster preparedness plans which include strategies for attaining access to remote populations; include HR requirements, also. –Thailand and Indonesia MAE 2005

	9. Explore with local NGOs possibilities for capacity sharing in disaster risk reduction. 10. Support capacity development for government at all levels for disaster risk reduction and disaster management. -Thailand and Indonesia MAE 2005
	11. Build vulnerability reduction into food security planning. 12. Trained "emergency core staff" should take a lead in developing (and periodically updating) the emergency preparedness plan (EPP) along with other staff. -WBG Strat Review 2004
	13. Lack of method in entry and exit strategies. -Ethiopia LL workshop 2003
	14. Planning framework should proactively address environmental concerns. -Kenya RAP 2001
	15. Exploit opportunities to work with the private sector as a business partner and as a development partner during emergencies. -South Africa LL 2003

	16. Refugee and other population movements are hard to predict; contingency plans should therefore be made for even the less likely scenarios. 17. CARE should nominate a person to ensure that regular analysis and scenario planning for the countries in the Balkans region is carried out. 18. CARE field staff should be briefed to observe the Red Cross CoC and the SPHERE standards. 19. CARE should consider producing a simple project checklist for program managers in emergencies. –Kosovo LLR
	20. We should develop a CO –specific contingency plan to avoid being caught off-guard in the future. –Sri Lanka AAR
	21. There is a need to develop emergency procedures and staff guidance for procurement. – Haiti 2005
	22. Emergencies personnel deployed to the field should have access to a 4x4 vehicle with driver and interpreter –West Bank/Gaza RTE 2002
Procurement and Logistics	1. Examine and assess procurement and logistics system in Sri Lanka. 2. Develop CARE and WV's capacity to handle procurement, warehousing and logistics into a common pool for faster response. –India and Sri Lanka MAE 2005
	3. Investigate possibilities for collaboration (among agencies) on procurement and capacity development for local marketing systems. –Indonesia and Thailand MAE 2005

	4. The purchase of materials has been most efficient where beneficiaries and CARE have made bulk purchases. 5. CARE WBG needs to make strategic decisions regarding its current deep involvement in procurement activities, related to questions about impact and appropriateness. 6. Longer-term planning and an exit strategy need to be designed in WBG re: procurement. –WBG Strat Review
	7. There is a need to select, stay with and insist on the utilization of a single commodity management system – either GIMS or Scala. –India earthquake 2001
	8. CARE should develop a permanent procurement capacity, located in Europe, for supporting future emergency response. 9. Emergency missions should have an overall coordinator who can supervise logistic procedures and ensure smooth links between procurement, transport of goods and warehousing. –Kosovo LLR
	10. Vendor lists for materials according to centers of excellence needs to be prepared and updated regularly.
	11. Huge increase in procurement needs means restructuring is necessary, not merely additional staff. 12. Pre-arrangement of transport is critical. –Indonesia/Aceh AAR

Finance	1. If time allows, use an external person to assess CO financial management and user requirements. -Iraq RTE II 2003
	2. Structured, formal procedures specific to emergencies could prevent disregard for normal procedures, such as purchases without purchase orders. -Haiti AAR 2005
	3. Ensure that funds are efficiently used. -Thailand and Indonesia MAE 2005
	4. Administration and finance staff should try and make regular visits to field offices. This is especially important for understanding operating constraints during emergencies. -West Bank/Gaza Strat Review 2004
	5. CARE should ensure that a finance manager is included at the start of all emergency operations. -Kosovo LLR
	6. There is a need to safeguard against overtaxing senior country office staff in trying to accommodate CARE information needs. -Afghanistan 2002
	7. A closer examination of current division of labor and staff efficiency / motivation would be useful in light of the increased workload resulting from disaster. -Haiti 2002
	13. If a CO lacks finance officers with appropriate experience, CARE should send a specialist for large-scale emergency operations. 14. CARE should ensure that CERT finance officers know the accounts software. -Afghanistan 2002
	15. There was a damaging mis- or under-utilization of standard CARE accounting procedures. -Haiti 2005

Psychology and staff wellness	1. Review and improve staff living conditions as necessary to a reasonable level. 2. Ensure that counseling services are available for returning staff that have been exposed to stressful situations. – Darfur RTE I 2004
	3. COs hosting a CERT should ensure adequate living and office arrangements for CERT that should facilitate team-building between the CERT and the CO. –Iraq RTE II 2003
	4. Affected staff had to wait 4 months to get psychosocial support. –Haiti AAR 2005
	5. Given the high potential for mental health disorders, consider immediate additional means of support. –Thailand and Indonesia MAE 2005
	6. When staff are under severe stress…management should pay close attention to morale, through encouragement and practical support. – 7. CARE should continue the good practice of providing professional counselors during emergencies. -Kosovo LLR
	8. Include psychosocial care for staff engaged in emergency response at all levels. –India Tsunami AAR

Communication	1. CI needs to define responsibilities and standards for maintaining an acceptable telecommunications policy. 2. Communications need to be recognized within CI as a priority both for security and operations; a minimum acceptable capacity needs to be defined. 3. Ensure staff are trained in the use of communications equipment and systems. --Iraq RTE II 2003
	4. Put in place a strategy to disseminate information up to the CI level. -Haiti AAR 2005
	5. CARE should ensure that it always has full communications from the start of an emergency. 6. Strict procedures for telephone use should be set up from the start of a mission.
Security	1. Having a security chain and functioning communication at the institutional level was essential for crisis management. -Haiti AAR 2005
	2. CARE should require all country offices to develop and update security management plans. 3. CARE should train security officers to understand their role. 4. Security decisions are best made by the staff in the field. - -Afghanistan 2002
	5. COs should ensure that security and safety plans are in place at the beginning of an emergency and are regularly updated. -Darfur RTE I 2004

	6. The issuing of ID cards and having a "full-fare, open return economy" air ticket are security issues and must be applied in forthcoming CERT deployments. –West Bank/Gaza RTE 2002
	7. CARE should ensure that all vehicles in emergencies are equipped with basic medical kits and equipment for dealing with vehicle accidents. –Kosovo LLR
Human Resources	1. Develop documentation to orient staff to intervention strategies such as food distribution. 2. Establish ToR before hiring any consultant or contractor. 3. In the future, re: international staff, we should identify genuine needs, define ToRs, and regularly update the human resources bank. –Haiti AAR 2005
	4. CARE International should train local staff in emergency mgmt skills, especially during lulls. 5. All staff deployed in an emergency should have a ToR agreed to in advance with the CO. CARE should develop standard ToRs for each of the CERT posts. 6. CARE must ensure that staff on the CERT roster are highly qualified and can add value to the response. 7. –Afghanistan 2002

	8. Articulate a more strategic approach to CERT skills selection and deployment in the CI Emergency Manual. 9. Clarify timeframe for CERT to make it more flexible and responsive to actual needs 10. Review and revision of generic ToRs for CERT members 11. Develop and implement a training program for CERT staff that includes coaching/mentoring techniques – Iraq RTE II 2003
	8. Matching criteria for selection of CERT personnel should be sensitive to the issue of 'nationalism' 9. CERTs must be prepared for almost total self-sufficiency. 10. COs must be capacitated to know what to expect of a CERT team if one is deployed to their country –West Bank/Gaza RTE 2002
	11. In all major disasters, deploy HR staff in the field offices in the early stages to enable local recruitment. 12. Ensure that in-country disaster preparedness capacity and plans are reinforced with a strong component of international and regional expertise so that COs are able to scale up response quickly by deploying optimum staff. –India and Sri Lanka MAE 2005

	13. Expand the collaboration potential for regularized multi-agency training in cross-agency areas of relevance. 14. Strengthen human resources policies to overcome barriers to human resource constraints and optimal performance –Thailand and Indonesia MAE 2005
	15. CARE should set up and train a Country Office core team in emergency response. –WBG Strat Review 2004
	16. Shadow roles and number and types of staff needed at any given time should be outlined before each emergency response. –Ethiopia LL 2000
	17. Early deployment priorities should include the rapid establishment of communications systems and setup of adequate accommodations for CARE staff. 18. CARE assessment teams should be gender balanced; PRA training would be a plus. –India Earthquake 2001
	19. Invest in staff and support systems at the onset of an emergency. –South Africa LL 2003
	20. When personnel change, a proper hand-over of responsibilities is essential, to ensure that the newcomer is fully briefed about the job and is able to benefit from the outgoing post-holder's knowledge, experience and professional contacts. 21. Mission Directors and human resource managers should ensure that all staff receive a job description. –Kosovo LLR

	22. Brief and train emergency officers and partner staff in participatory and gender analysis tools. 23. Recruit female staff and volunteers for assessments and other responses. 24. Performance management of the deployed staff needs to be incorporated within the Annual Performance Appraisal process. Reward and recognition of deployed staff needs to be developed. -India Tsunami AAR
	25. Use generalists with flexibility to fill gaps. 26. Strong HR needed at all levels. -Indonesia and Aceh Tsunami AAR
	27. HR should be involved in the proposal development stage to assist in staff recruitment. 28. HR should prepare an in-house list of staff expertise available and develop an emergency roster. -Garowe AAR

	29. CARE Sri Lanka HR should make regular visits to field offices to communicate change processes and listen to staff concerns. 30. We need to find ways to build local HR capacity and facilitate regional recruitment. 31. HR to monitor conditions of inequity and ensure staff concerns are heard. 32. CI should maintain an active roster of CVs of people who can be called upon when an emergency arises. 33. CARE USA should dedicate a full-time person to emergency recruitment and placement in emergency situations. –Sri Lanka AAR
	34. Better define the profiles and competencies needed for personnel engaged in emergency operations, including data management officers and warehouse managers. 35. Train CARE personnel in EPP and existing protocols --Niamey AAR 2005
	36. At least one senior decision-maker should visit at the beginning of an emergency operation to aid communication. –Afghanistan 2002
	37. Senior staff should visit crisis areas early and regularly. –Darfur RTE 2004
	38. We need to build capacity in the field to decentralize to regional offices by enhancing the authority level. –Sri Lanka AAR

	39. Visits by senior management should take place as close to the beginning of an emergency as possible. Iraq RTE II 2003
Information Management	40. Apply new information management procedures as soon as possible. -West Bank/Gaza RTE 2002
	41. Info management is critical in an emergency and must be developed prior to a disaster. -India Earthquake 2001
	1. Establish MIS inventory to identify what type of information is already being collected and how, and what information is/is not available. -South Africa LL 2003
	2. CARE, UN agencies, donors and govt need to establish clear reporting lines, managed by a centralized body using a standard format. -South Africa LL 2003
Policy and Advocacy	3. Contingency plan should address what advocacy issues CARE should raise. 4. CERT should deploy an advocacy specialist early in an emergency response if there is not an individual with this expertise on the ground already. -Afghanistan 2002

	5. Critical capacities, such as advocacy and policy analysis, should not be attached to ad hoc structures. 6. Add an advocacy and policy analysis "box" to the aCERT organigram and identify a focal point from within CI membership to take on responsibility for developing and maintaining the required capacity. 7. CARE should aim to develop a more focused strategy on advocacy during emergencies via a CO or CERT focal point. 8. There should be an understanding that the preparation of scripts for congressional testimonies or other high-level representation be done in close consultation with the CO. --Iraq RTE II 2003
	1. Devote additional resources to advocacy. 2. Advocacy for land rights. -Thailand and Indonesia MAE 2005
	3. Advocacy on the right to access to adequate quality and quantity of water 4. Advocacy should be professional, should be based on our own work and direct experience. 5. Legal advice in terms of RBA is needed -WBG Strat Review 2004
	6. A key CARE role should be advocacy for disaster victims to inform them about their rights for relief and rehabilitation. -India earthquake 2001
	7. CARE should agree on a global policy defining its relationship with the military. -Kosovo LLR

	8. Appeals and SitReps were important advocacy tools with donors and lead CARE bodies; the appeal mechanism should be institutionalized. -India Tsunami AAR
	9. Develop an advocacy strategy for the emergency response. –Niamey AAR 2005
Decision Making	10. CARE Lead Members should state the lines of authority at the beginning of an operation. –Afghanistan 2002
	11. The CO should have an emergency contingency plan with clear responsibilities and lines of communication laid out. –Haiti 2005
	1. Develop training program for CERT staff that includes exposure to a variety of operations within various COs, especially different financial and administrative systems. 2. A senior CI staff representative should facilitate adjustments and/or clarifications of roles and responsibilities of CO and CERT staff. -Iraq RTE II 2003
	3. Plan a meeting of all program staff immediately after a disaster to delegate and clarify each person's tasks. –Haiti AAR 2005
	4. Set up a contingency emergency plan that is pyramidal; ensure communication about the new structure and involve the community in decision-making. 5. At the beginning, lines of responsibility and leadership for managing the emergency were not well established, leading to frustration and delays. –Haiti AAR 2005

	6. Reporting lines within a CO should be re-articulated as soon as possible after arrival of external emergency response personnel –West Bank/Gaza RTE 2002
	7. Ensure senior managers in CO and districts take responsibility for and participate in all coordination fora – India and Sri Lanka MAE 2005
	8. We need a common position…the problem is to know who does and decides what in the CARE network. We need someone to prepare (an advocacy) paper, circulate it, and get agreement. –WBG Strat Review 2004
	9. Roles and responsibilities of representatives should be clearly defined before emergency ops to promote clear understanding of roles; validate at time of emergency. –Ethiopia LL 2000
	10. CARE, UN agencies, donors and govt need to establish clear reporting lines, managed by a centralized body using a standard format. –South Africa LL 2003
	11. Lead or Temporary Coordinating Member and their country offices should take responsibility for monitoring potential emergencies; CARE should decide where responsibility lies for analysis and contingency planning when there is no Lead or Coordinating member. 12. Project managers should be briefed on all their responsibilities, and preferably equipped with a checklist of these. –Kosovo LLR

	13. Clarify decision making structure. 14. There should be proper authority delegation to make decisions at the field level. -Garowe AAR
	15. Delegation of authority to field offices in the early phases has been empowering for staff. We need to ensure the delegation of authority is systematized. -Sri Lanka AAR
	16. Define and clarify the roles and responsibilities, in terms of preparation, response, communication, and training, of CO, CARE USA, CEG, and CI. -Niamey AAR 2005
Evaluation and Learning	1. CARE International should include application of lessons previously learned in future ToRs. 2. CARE should audit emergency operations within 3 to 4 months of the start of the operation. -Afghanistan 2002
	3. CARE should make use of the Sudan experience to develop a "good practice" guide for humanitarian protection activities. 4. Ensure that adequate resources for learning are built into project budgets to improve timeliness and effectiveness of M&E events. --Darfur RTE I 2004
	1. Encourage COs to allocate adequate resources in project budgets to cover costs for lessons-learned sessions and M&E activities. - Iraq RTE II 2003

	2. We should promote a culture of continuous reflection and willingness to make modifications… 3. We should identify a focal point for early collection, processing and analysis of data. 4. SPHERE standards should be adopted and applied in the future. -Haiti AAR 2005
	5. How can CARE disseminate and share learning from successful shelter work in Sri Lanka? 6. Ongoing monitoring of watsan data in India will be important. –Sri Lanka/India MAE 2005
	7. The Inter-Agency Working Group should look into putting monitoring expertise on the ground in the early stages of an emergency to assist with establishing M&E systems. –Thailand and Indonesia MAE 2005
	8. Improve the level of household-level qualitative analysis. 9. Improve monitoring o f impact in projects by including indicators, peer and inter-agency reviews, etc., and staff training in these methods. –WBG Strat Review 2004

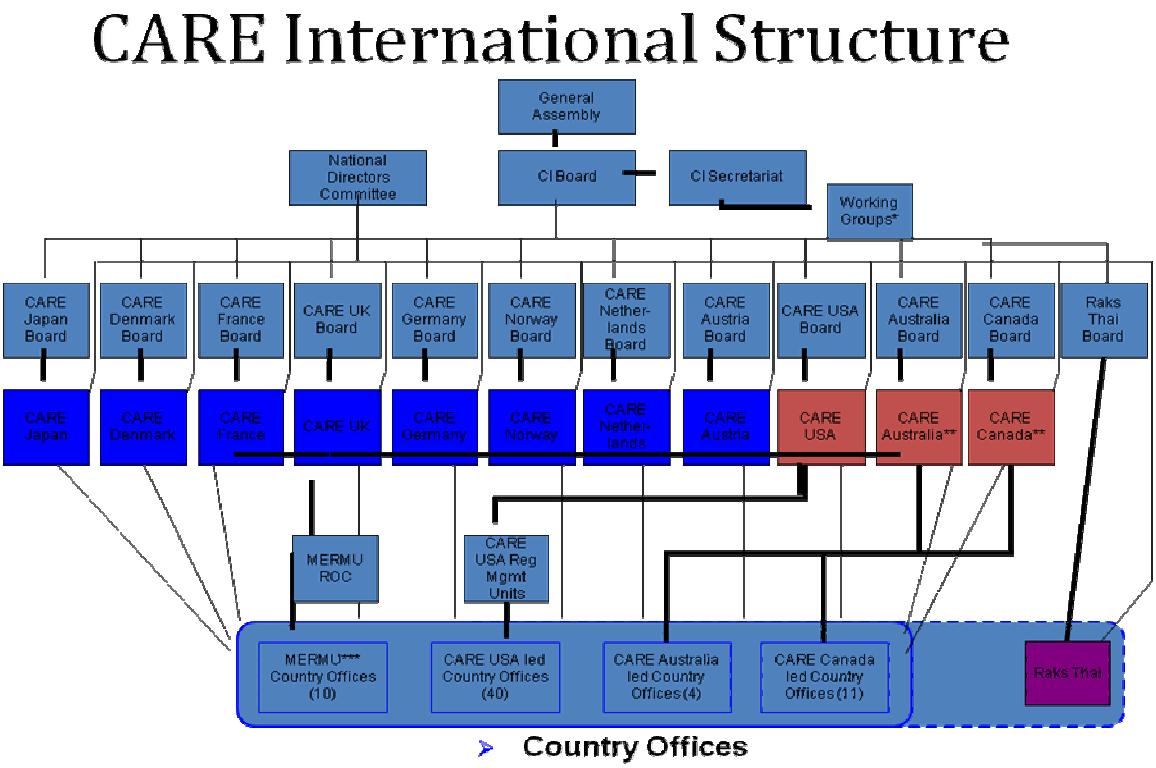
	10. Need for methodology / clear guideline for monitoring food distribution activities. 11. During impact evaluation, CARE should study whether beneficiaries felt that the food aid they received was appropriate or not. 12. Absence of guidelines for evaluating emergency response limited initial learning and assessment of impact and further impacts design of future interventions –Ethiopia LL 2000
	13. Assessment of needs should also include the assessment of local capacities and coping mechanisms to inform and guide external interventions. – India earthquake 2001
	14. Establish a DM&E unit and a DM&E coordinator at a CARE CO level, as well as M&E Manager at a project level. 15. Establish standard formats for M&E systems across projects. 16. Project budgets should include the cost of a full time DM&E manager and shared costs of DM&E coordinator. 17. Include field-based learning events for CARE and partners in project design and budget –South Africa LL 2003

	18. CARE needs to insist, with donors and amongst its members, on rigorous assessment before committing itself to an emergency program. 19. CARE projects need to be accompanied by an effective system that regularly assesses the impact of its activities in beneficiaries. 20. CARE members should do their own internal evaluations as to which sectors of their recruitment performed well and which less well, with a view to making improvements as necessary. – Kosovo LLR
	21. Collect and solicit data disaggregated by sex to understand composition of surviving families. 22. Assessment reports should reflect the needs for different phases. 23. Capture the lessons learnt in the current and previous disasters to form a base for future programming. 24. Monitoring should include impact indicators. –India tsunami AAR
	25. Thus far we are consistently monitoring at the output level only. We should more consistently collect and analyze quality data. 26. Develop M&E framework / strategy for working in emergencies. -Sri Lanka AAR

	27. There should be a formal process of sharing experiences with other organizations working on the tsunami response. 28. Clear achievement indicators and reporting formats are needed -Thailand AAR
	29. Establish and reinforce monitoring and evaluation in emergency plans. 30. Develop a working group that will assure the utilization of lessons learned and the development of a nutrition strategy for crisis and non-crisis periods. -Niamey AAR 2005

APPENDIX C

CARE Organigram



APPENDIX D

Interview Protocol, Use of Evaluations

1. From your perspective, what are the critical events where planning takes place?

    a. What are the major meetings, workshops, retreats, etc., where planning takes place *for your job*?

    b. From your perspective, at what key events does planning take place *for CARE as a whole*?

    c. Is most planning done in a routine, regular fashion, or are there ad hoc meetings where major planning occurs? Explain.

2. Who are the main participants for the key planning events you described?

    a. Who organizes the events?

    b. Who facilitates the events?

    c. Who determines the content of the events?

d. Who is present for the events?

3. Where do these events tend to occur? How long do they last?

4. During these key meetings, routine or ad-hoc, has the discussion included emergency response or capacity building?

a. On a policy level?

i. If so, what was the context?

ii. If so, were any decisions made on a policy level?

b. On an operations level?

i. If so, what was the context?

ii. If so, were any decisions made on an operations level?

c. When it is necessary to decide between several courses of action, how do you prioritize what to focus on?

5. Are evaluation findings an information source for these planning events?

a. do you regularly receive or have easy access to evaluation findings from disasters and emergencies? Where would you go to find such information if you wanted it?

b. Are "lessons learned" incorporated into discussions about the future at major planning events?

c. Do certain "lessons" or types of lessons get used more than others? If so, why and how?

d. Have you been involved in any disaster or emergency-related evaluations?

i. If so, what was the context and what was your role?

ii. If so, were you enthusiastic about the impending evaluation, or were you reluctant? Why?

6. Are lessons-learned being used appropriately and to effect for planning? If not, how could this be improved?

a. do the appropriate people have access to lessons-learned?

b. are lessons-learned put in a format that is easy to use and understand? If not, what would help?

c. are lessons learned received in a format that is easy to read and use? If not, how would you change the format?

d. is the timing of the delivery of lessons-learned such that they are immediately useful? If not, how could timing be improved?

7. Do you have other remarks or suggestions for improving the use of evaluation reports?

REFERENCES

Alkin, M. C. (1985). *A Guide for Evaluation Decision-makers*. Thousand Oaks: Sage.

Alkin, M. C., & Taut, S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation, 29*, 1-12.

Almeida, C., & Báscolo, E. (2006). Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cad Saude Publica, 22*, S7-S19.

Argyris, C., & Schon, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison Wesley.

Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison Wesley.

Berg, E. (2000). Why Aren't Aid Organizations Better Learners? In J. Carlsson & L. Wohlegemuth (Eds.), *Learning in Development Co-Operation* (pp. 24-40). Stockholm: SIDA.

Beyer, J. M., & Trice, H. M. (1982). The Utilization Process: A Conceptual Framework and Synthesis of Empirical Findings. *Administrative Science Quarterly, 27*(4), 31.

Cahill, K. M., M.D. (Ed.). (2003). *Basics of International Humanitarian Missions*. New York: Fordham University Press.

Chelimsky, E. (1997). The Coming Transformations in Evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st Century: A Handbook*. Thousand Oaks: Sage.

Conner, R. F. (1998). Toward a Social Ecological View of Evaluation Use. *American Journal of Evaluation, 19*(2), 237-242.

Cooley, A., & Ron, J. (2002). The NGO Scramble: Organizational Insecurity and the Political Economy of Transnational Action. *International Security, 27*(1), 34.

Dichter, T. (1999). Globalisation and its Effect on NGOs: Efflorescence or Blurring of Roles and Relevance? *Nonprofit and Voluntary Sector Quarterly, 28*(supplement 1999), 55-56.

Donahue, A. K., & Tuohy, R. V. (2007). Lessons We Don't Learn: A Study of the Lessons of Disasters, Why We Repeat Them, and How We Can Learn Them. *Homeland Security Affairs, 3*(3).

Dufour, C., Véronique Geoffroy, Hugues Maury, and François Grunewald. (2004). Rights, Standards, and Quality in a Complex Humanitarian Space: Is SPHERE the Right Tool? *Disasters, 28*(2), 124-141.

Ebrahim, A. (2003a). Accountability in Practice: Mechanisms for NGOs. *World Development, 31*(5), 813-829.

Ebrahim, A. (2003b). Making Sense of Accountability: Conceptual Perspectives for Northern and Southern Nonprofits. *Nonprofit Management and Leadership, 14*(2), 191-212.

Ebrahim, A. (2005). Accountability Myopia: Losing Sight of Organizational Learning. *Nonprofit and Voluntary Sector Quarterly, 34*(1), 56-87.

Edwards, M., & Hulme, D. (1996). Too Close for Comfort? The Impact of Official Aid on Nongovernmental Organizations. *World Development, 24*, 961-973.

Forss, K., Cracknell, B., & Samset, K. (1994). Can Evaluation Help an Organization to Learn? *Evaluation Review, 18*(5), 574-591.

Fowler, A. (1997). *Striking a Balance*. London: Earthscan.

Frerks, G., & Hilhorst, D. (2002). *Evaluation of Humanitarian Assistance in Emergency Situations*: UNHCR.

Gibelman, M., & Gelman, S. R. (2001). Very Public Scandals: Nongovernmental Organizations in Trouble. *Voluntas: International Journal of Voluntary and Nonprofit Organizations, 12*(1), 49-66.

Ginsburg, A., & Rhett, N. (2003). Building a Better Body of Evidence: New Opportunities to Strengthen Evaluation Utilization. *American Journal of Evaluation, 24*(4), 489-498.

Hallam, A. (1998). *Evaluating Humanitarian Assistance Programmes in Complex Emergencies*. London: ODI.

Henry, G. T., & Mark, M. M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation, 24*(3), 293-314.

Hilhorst, D. (2002). Being Good at Doing Good? Quality and Accountability of Humanitarian NGOs. *Disasters, 26*(3), 193-212.

Kirkby, J., Howorth, C., Keefe, P., & Collins, A. (2001). A Survey of Evaluation Experience in Complex Emergencies. *The International Journal of Human Rights, 5*, 114-129.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions in Evaluation, 2000*(88), 5-23.

Leviton, L. C. (2003). Evaluation Use: Advances, Challenges, and Applications. *American Journal of Evaluation, 24*(4), 10.

Mark, M. M., & Henry, G. T. (2004). The Mechanisms and Outcomes of Evaluation Influence. *Evaluation, 10*(1), 35-57.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco: Jossey-Bass.

Merriam-Webster Online Dictionary. (2007). Retrieved January 4, 2008, from http://www.m-w.com

Nietzsche, F. (1883). Thus Spoke Zarathustra. In G. Colli & M. Montinari (Eds.), (Vol. 4).

O'Donnell, M. R., D. Bacos, and M.L. Bennish. (2002). Nutritional response to the 1998 Bangladesh Flood Disaster: Sphere Minimum Standards in Disaster Response. *Disasters, 26*(3), 229-241.

Patton, M. (2002). Utilization-focused Evaluation. In *Evaluation Models* (pp. 425-438).

Patton, M. Q. (1988). The Evaluator's Responsibility for Utilization. *Evaluation Practice, 9*(3).

Patton, M. Q. (1994). Developmental Evaluation. *American Journal of Evaluation, 15*(3), 311-319.

Patton, M. Q. (1998). Discovering Process Use. *Evaluation, 4*(2).

The Program Evaluation Standards. (1994). (2nd ed.): Sage Publications, Inc.

Ramalingam, B. (2005). Implementing Knowledge Strategies: Lessons from International Development Agencies [Electronic Version]. *ODI Working Paper, 244*. Retrieved August 07, 2007,

Sackmann, S. A. (1992). Culture and Subcultures: An Analysis of Organizational Knowledge. *Administrative Science Quarterly, 37*(1).

Sandison, P. (2006). *Evaluation Utilisation*. London: ODIo. Document Number)

Scriven. (1991). *Evaluation Thesaurus* (4th ed.). Thousand Oaks: Sage.

Senge, P. M. (1992). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Century Doubleday.

Shulha, L. M., & Cousins, J. B. (1997). Evaluation Use: Theory, Research, and Practice Since 1986. *American Journal of Evaluation, 18*(1), 195-208.

Shulock, N. (2000). The Paradox of Policy Analysis: If It Is Not Used, Why Do We Produce So Much Of It? *Journal of Policy Analysis and Management, 19*(2).

Simmons, P. J. (1998). Learning to Live with NGOs. *Foreign Policy, 112*, 82-96.

Smillie, I. (1997). NGOs and development assistance: a change in mind-set? *Third World Quarterly, 18*, 563-578.

Smillie, I., and Larry Minear. (2004). *The Charity of Nations: Humanitarian Action in a Calculating World*. Bloomfield, CT: Kumarian Press.

Smith, N. L. (1989). The Weiss-Patton Debate: Illumination of the Fundamental Concerns. *American Journal of Evaluation, 10*(1), 5-13.

Weiss, C. H. (1977). *Using social research in public policy making*. Lexington, Mass.: Lexington Books.

Weiss, C. H. (1988). If Program Decisions Hinged Only on Information: A Response to Patton. *Evaluation Practice, 9*(3),13.

Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation, 19*(1), 21-33.

Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. *American Journal of Evaluation, 26*(1), 12-30.

Willits-King, B., & Darcy, J. (2005). *Review of CARE's Emergency Preparedness and Response Capacity*. London: Overseas Development Institute.

Wood, A., Raymond Apthorpe, and John Borton (Eds.) (2001). *Evaluating International Humanitarian Action: Reflections from Practitioners*. London: Zed Books.

VITA

Monica LaBelle Oliver was born in Seattle, WA, in July of 1974. She received a B.A. in Independent Studies in the Romance Languages from Loyola Marymount University in May of 1996. She served in the Jesuit Volunteer Corps in Bordeaux, France, from 1996-1997. After working briefly for a water purification corporation in Los Angeles, she completed coursework toward a Master's degree in Public Administration / International Development from 1999-2000 at Rutgers University-Camden. She served in the Peace Corps in Mali, West Africa, from 2000-2002, working as a small enterprise development specialist with a microenterprise INGO. Her master's thesis, completed in 2002, evaluated that INGO's microenterprise program in the context of similar programs in Mali. She has resided in GA since 2003, while working on a doctorate in Public Policy from Georgia State University and Georgia Institute of Technology, with an emphasis in Policy and Program Evaluation. She presently lives with her husband, Erik, and her son, Lucas, in Oxford, GA.