

What We Know About Joint Evaluations of Humanitarian Action

Learning from NGO Experiences

Section 2 of 3: THE STORIES

April 2011

Contents

ACKNOWLEDGEMENTS.....	3
ABOUT THIS BOOKLET	4
THE STORIES	5
Niger	5
Guatemala	8
Indonesia	11
References and Further Reading	14

ACKNOWLEDGEMENTS

Many people have shared their valuable experience and time in the creation of this booklet. Special thanks go to all of them, particularly, in Guatemala, Carla Aguilar of Save the Children US, Hugh Aprile of Catholic Relief Services, Borys Chinchilla of Mercy Corps, and Juan Manuel Giron Durini of the ECB Project; in Niger, Jasmine Bates and Marianna Hensley of Catholic Relief Services, and Julianna White, of CARE; in Indonesia, Adhong Ramadhan and Josephine Wijastuti of Catholic Relief Services, Agus Budiarto and Evi Esaly Kaban of Save the Children, Harining Mardjuki and Anwar Hadipriyanto of CARE, and Richardus Indra Gunawan and Yacobus Runtuwene of World Vision International. Special thanks also go to John Wilding, Pauline Wilson, John Telford, Maurice Herson of ALNAP, Jock Baker of CARE and Guy Sharrock of Catholic Relief Services who have given critical input into this work. Malaika Wright was the author of the first paper.

The April 2011 version of the booklet was updated by Katy Love from the ECB Project, Loretta Ishida of Catholic Relief Services, Jock Baker of CARE, Hana Crowe of Save the Children, and Kevin Savage of World Vision. The booklet was revised based on feedback and reports from those who participated in ECB-supported joint evaluations in 2010 in Indonesia, Haiti, the Horn of Africa, and Niger. These people served as evaluation managers and coordinators, team leaders, team members, Steering Committee members, ECB field facilitators, and ECB accountability Advisers in joint evaluations, including: Paul O'Hagan, Greg Brady, Yves-Laurent Regis, Angela Rouse, Katy Love, and Jock Baker (Haiti); Yenni Suryani, Pauline Wilson, Loretta Ishida, and LeAnn Hager (Indonesia); Kevin Savage, Chele DeGruccio, Jim Ashman, and Wynn Flaten (Horn of Africa); and Kadida Mambo (Niger). The ECB Project thanks all who contributed to this work.

ABOUT THIS BOOKLET

This booklet was written to share knowledge gained from the experiences of people that have been involved in joint evaluations conducted by non-governmental organizations (NGOs). It mainly profiles the work of NGOs involved in the Emergency Capacity Building Project (ECB), which has a goal to improve the speed, quality and effectiveness with which the humanitarian community saves lives, improves the welfare, and protects the rights of women, men and children affected by emergencies.

This booklet also draws on the lessons of multi-agency evaluations that already exist within the humanitarian sector. Major contributions have come, in particular, from the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP).

We hope that learning from previous experiences captured here will be useful for all those considering leading their agencies through a joint evaluation. The learning shared here is targeted at evaluation practitioners, managers, and NGOs contemplating a joint evaluation. Additionally, we hope that it will contribute to a growing body of knowledge on these processes and show that while there are many unanswered questions about joint evaluations, there is a lot we already know.

The booklet is comprised of three sections. The first section, **The Guide**, can be referred to as a ‘how-to’ for those closely involved in joint evaluations. It provides a framework for those approaching an interagency evaluation. This section, **The Stories**, shares several case studies from the ECB Project’s experiences. The third section, **The Tools**, includes many templates and tools that can be adapted for evaluations, including sample terms of references, agreement documents, and checklists.

THE STORIES

Niger

About the *Joint Independent Evaluation of the Humanitarian Response of CARE, Catholic Relief Services, Save the Children and World Vision to the 2005 Food Crisis in the Republic of Niger*⁷

A crisis

When reports of Nigerien hunger, poor harvests, sick frail children and a slow humanitarian response finally caught the world’s attention, it was too late for many. By the time the UN received its first pledge in July 2005, and grain began its slow voyage across the ocean from India, the window for pre-emptive action in Niger had closed. NGOs on the ground tried to help as many people as they could. CARE, Save the Children, Catholic Relief Services, and World Vision mounted responses that reached nearly 1.5 million people in the first round of food distributions, or 51 percent of the vulnerable.ⁱ

However, it seemed that the humanitarian community as a whole had missed the mark. Controversy over the emergency in Niger heated up in July and August of 2005. Why had the response been so slow? Had early warning systems failed? Did the media sensationalize the emergency? And was the Nigerien president’s defensive posturing an attempt to cover up the extent of the crisis?

Why a joint evaluation?

While conducting an emergency assessment for CARE Niger in August, Jock Baker, a program quality coordinator from CARE, felt that Niger was a good candidate for a joint evaluation. Having been involved in a multi-agency evaluation of the Tsunami response conducted by CARE, World Vision and Oxfam, Baker felt that such an approach would work in Niger. “It was a big area and there seemed to be a systemic failure as a whole of the humanitarian community,” he said, reasoning that a multi-agency evaluation would help fit the different pieces of the puzzle together.

He approached his colleague, Amadou Sayo, the assistant country director of CARE Niger, with the idea. Amadou was enthusiastic and promised to contact his colleagues at other agencies to check their interest. The idea took off and Save, World Vision and Oxfam agreed to partner with CARE on the evaluation and assign steering committee members.

In October, with food distributions still ongoing, the four agencies commissioned a joint evaluation of their responses. The evaluation yielded rich lessons not only about the crisis and its response, but also about the process by which such joint exercises are conducted.

The best and the worst of timing

One such lesson concerns the timing of the evaluation. Marianna Hensley, a program quality director for Catholic Relief Services, took over from a colleague as a steering committee member for the evaluation. When she joined the steering committee, about three weeks after the inception of the joint evaluation, they were reviewing the Terms of Reference for the evaluation team members to be hired. At the time of

⁷ This report can be viewed at www.alnap.org

evaluation she was grateful for 12-hour days. Upwards of 14 hours per day was the norm for Marianna and her colleagues during the ending phase of CRS’ food distribution in Niger. She was helping a partner organization set up its therapeutic feeding program, managing nine budgets and their related activities, and communicating with three field offices. Things were even more intense for field staff who were in the field for weeks on end without a break. All were exhausted. Yet as the end of food distributions drew ever closer, it merely promised the start of recovery activities.

“We were in the middle of emergency operations, particularly food distribution. Had [the evaluation] not been pushed from the outside, I don’t think it would have been done,” she said.

Still, Marianna concedes, there were some merits to the timing of the evaluation. “A lot of the focus of the Niger multi-agency evaluation was the extent of the food distribution and [to some extent] on the supplemental and therapeutic feeding. So theoretically, timing was right to come in at the end of the distribution to get the freshest perspectives on what happened.”

Whose idea was this?

Another lesson was that prior experience with or guidance about joint evaluations is necessary. The four country offices had never done a joint evaluation before and no one had time to give it much thought: “There wasn’t a whole lot of reflection that went into the process,” said Marianna. “We were told, ‘Here’s a Terms of Reference. Go participate in it.’ Steering committee members went to meetings, looked around at each other, and no one seemed to know what the original vision for deciding on a joint evaluation had been.”

Breaking the news

Yet another lesson concerned the presentation of the rather surprising results. When the team presented the findings for the first time, the lead evaluator had the unenviable task of explaining to four agencies that in responding to a famine, they had not quite understood the true nature of the crisis and that their response, though effective, had been too late (through no fault of theirs) to help avert the crisis.

However, one observer notes that it was very helpful for the same findings to be shared with several agencies represented in one room. “[The multi-agency evaluation] created a globally accepted version of events in Niger particularly about how things came about. Initially there were many versions/reasons as to why the long term crisis developed as it did.”

A Surprising Discovery

In July 2005, before he became involved in the evaluation, people were coming to John Wilding and asking for his interpretation of what was going on in Niger. But far away from Niger, the food security expert was as baffled as they were. All he could say was that he didn’t know. Like them, he was simply watching it unfold on television.

Even after he accepted the assignment as a team leader for the CARE, CRS, Save and World Vision multi-agency evaluation and arrived in Niger with the rest of the team, the answer was not any more apparent. “When we got there, we had team meetings at 6 o’ clock every night for a week, asking ‘Is this a crop failure? A famine?’ We didn’t know.” One of the team members had a background in marketing. Wilding gave her a car and asked her to go and find out what had happened from a market point of view. “She was a street fighter,” he said. “She went to the small merchants and the big merchants and she asked ‘What has happened? [You know,] we’re all in this for money.’ And they told her.”

The conclusion that the team arrived at was not a popular one. It was not one that the humanitarian community and even universities were talking about. Certainly not the media, which was portraying it as a famine. The word from the street was that it wasn’t a famine. It was a case of localized crop failures. “I think the government was correct to keep it under wraps,” said John. “I think the government understood what the problem was and I believe outsiders created a panic.” During the course of the evaluation, the people they talked to further confirmed this conclusion.

Hindsight

Those who were involved in the evaluation believe that despite the difficulties of getting everybody involved in the process, the evaluation proceeded more or less smoothly, and the team found plenty of people with much to say.

Having gained a common understanding of the Niger crisis and established working relationships, the partner agencies continued to meet long after the evaluation team had left Niger. They formed an NGO coordination forum, called the GDCI, and invited other agencies to join. The Office for the Coordination of Humanitarian Affairs began to come to their meetings. They submitted a joint paper to the government of Niger describing the status of the GDCI and other coordination bodies in Niger and successfully lobbied the World Food Program for blanket feedings in vulnerable areas in early 2006.

In June of 2006, with some support from the ECB Project, the GDCI convened a workshop to chart a work plan and review progress against the findings of the evaluation.

Guatemala

About the *Multi-Agency Evaluation of the Response to the Emergency Created By Tropical Storm Stan in Guatemala – CARE, Catholic Relief Services, Oxfam*⁸

An unusual opportunity

In late 2005, an unusual opportunity presented itself. The place was Guatemala, where half a million people were still reeling from the effects of Hurricane Stan. Six out of the seven ECB agencies were present in Guatemala and had been active in the response. They had shared information and coordinated their efforts thanks to their ongoing work in ECB's Disaster Risk Reduction initiative. Given the sudden nature of the emergency, it had been an unanticipated coordination exercise and the agencies were keen to glean some lessons from the event. Members of ECB's Accountability and Impact measurement initiative also saw this as an opportunity to further their objectives and urged the agencies to consider doing a joint evaluation of their response with some ECB support.

In December, emails flew back and forth between staff from the various agencies at headquarters and in the field to see if the agencies would make a commitment to do a multi-agency evaluation. ECB was pressing for a decision as time was marching on and good evaluators are hard to find at short notice.

Advocates for the evaluation among the six agencies in Guatemala felt that the evaluation would give them a good baseline for measuring the future effectiveness of their work. Some of the agencies were considering evaluations anyway. They also felt doing the evaluation together would help them optimize resources. However, some of the agencies were skittish, not wanting to look bad.

Finally, all the talk led to an agreement that the joint evaluation and case study on coordination would be combined under the leadership of one external consultant and that work would start in late January. The evaluation steering committee was formed, met, and amended the joint evaluation terms of reference (ToR) accordingly to include objectives to review coordination. Unfortunately, the agencies couldn't agree on whether to study impact or process. With these changes, parts of the ToR were not so clear. Some of this was due to translation as the English version of the ToR was translated into Spanish, the revised ToR translated from Spanish to English, and so on. In addition, local ownership was low as the ToR had come from headquarters. Agencies in Guatemala said that it would have been better to have started with a clean sheet of paper.

A non-Spanish speaker probes unanswered questions

By the time January rolled around, there were many unanswered questions: *Who was recruiting the other evaluation team members (three more were needed)? How were comments to be integrated into the ToR? What was the budget and how were costs to be shared?* Given these unknowns, Pauline Wilson, the ECB accountability and impact measurement project manager, decided to go to Guatemala in early January even though her Spanish speaking skills were nil.

The visit was instructive. Not all agencies were clear on what the evaluation would really offer them and had not reached consensus on what to accomplish through the evaluation. The steering committee agreed that Wilson should meet with each agency director individually and see whether they wanted their agency to participate. It was agreed that if three agencies were interested in evaluating their emergency response activities, the evaluation would go ahead.

⁸ This report can be viewed at www.alnap.org

A series of individual meetings with each agency director ensued. Three agency directors agreed to evaluate their emergency response. All six said they would be interested in learning more about the coordination effort that went on and ECB's effectiveness in communication and information gathering on behalf of the agencies. Views from the individual discussions with directors were shared during an ECB steering committee meeting along with each agency's priority objectives, budget and cost sharing information and a schedule for the team.

A confusing Terms of Reference

The above discussions resulted in yet more changes to the ToR to describe the different ways that agencies wanted to participate. It became even more complicated as it was translated back and forth from Spanish to English and new changes were not reconciled clearly enough. The steering committee assumed that the team leader would meet with the agency directors or their representative soon after their arrival in Guatemala and negotiate what was doable, and clarify what they meant by the many words that were used in the ToR. This is a common practice for most evaluations, but it never happened for this one.

Who does what?

A major challenge in doing joint evaluations is getting a commitment from one agency office in a country to coordinate the activity on behalf of the other agencies. The coordinating office is expected to oversee the recruitment of the evaluation team and ensure they have support, e.g. a workspace, transport, per diem, help them organize meetings between the agencies, etc. These responsibilities were dispersed across a number of offices; a long list of responsibilities was put on a flipchart and people did put their name next to each specific responsibility.

One agency did commit to being responsible for recruiting the evaluation team and overseeing much of the evaluation process, though the person doing the work left just as the evaluation team began the evaluation in February. Her transition out at such a crucial moment affected the continuity of the process and left the evaluation team without guidance on where to go for support

While not having a clear lead presented many problems, agencies in Guatemala considered this sharing of tasks an interesting coordination effort, which pressured them to share responsibilities for a common goal.

In the field

With no briefing upon arrival and no clear lead agency to go to, the evaluation team lacked clarity and guidance from the agencies in Guatemala on what the evaluation was to achieve. The evaluation team decided to define their own field visit plan and almost created its own evaluation purposes. The team chose a very wide field sample, with a very tight schedule and a strong quantitative approach that in the end made it difficult to have a comprehensive understanding of the field reality. In some cases too much data was collected, yet some essential data to understand the impact agencies made in the field, especially from a beneficiary perspective, was missing. Agencies also did not convey to the evaluation team a clear understanding of what they meant by impact.

A "critical" presentation

The evaluation started in February and ended in late April. None of the agencies were happy with the results when they were presented. It wasn't just the quality of the information being presented and that it

did not help them understand the impact. At least one of the agencies felt that the presentation was excessively critical and made unfair comparisons, e.g. comparing one agency against another in areas where they had very different types of interventions.

But most of the factors leading to the evaluation's undoing had happened long before the presentation of the findings. Lack of clarity about roles and responsibilities, a poor recruitment process for the evaluation team, unclear objectives, and lack of understanding of joint evaluations are only some of the reasons why the evaluation process and report did not adequately meet the overall objectives.

After the dust settled

Disappointed, the agencies in Guatemala nonetheless took the findings seriously and each carried out an internal review process. ECB observers of the evaluation chalked up the Guatemala evaluation as a case of the "what not to do's" in joint evaluations. However, when another member of the ECB team visited Guatemala in October 2006, she found some surprising perspectives on the evaluation.

For one thing, the agencies were planning a lessons learned workshop about the evaluation. At a meeting of the country directors of the agencies that same month, they had referenced the evaluation and how useful it had been. Even the agency that had been least happy with the results had made some improvements based on the recommendations, such as integrating risk reduction into its regular programming and adopting emergency protocols on where it should and shouldn't respond. It had also begun operations in a new area of work, thanks in part to what it learned from its partners about the work in that area. Another agency realized the importance of staff training, had its staff participate in Sphere training, and has plans for future training on needs assessment.

One country director believes his agency learned a lot regarding following the Sphere standards, working with partners and working with the local government, in part because of the opportunity to learn from the programs and experiences of the partners. Additionally, he claims the agencies learned the importance of sharing not only information but tools, so as to avoid duplication. After the evaluation, the agencies took some steps to standardize things like assessment tools.

The assistant country director of one agency concedes that the evaluation was useful for his agency in that it validated some of their work and provided constructive criticism. However, "it wasn't in-depth enough to help us in our programmatic response," he concluded.

"Initially, this process had a lot of inconveniences, but at the end we have a lot of richness," said one of the staff who had been involved in the evaluation.

Indonesia

About the CARE, Catholic Relief Services, Save the Children and World Vision Indonesia Joint Evaluation of their Responses to the Yogyakarta Earthquake

In many communities you could still see the rubble of houses that used to be and telltale cracks in structures still standing. But recovery was well underway in affected areas of Yogyakarta, Indonesia when an evaluation team for Catholic Relief Services (CRS), Save the Children UK, World Vision and CARE visited communities a year after the earthquake that killed over 5,000 in May 2006. Just how effective the agencies had been in assisting those communities with recovery was what the team was there to discover. For the ECB Project, which had provided support to previous evaluations in Niger and Guatemala, there was also much interest in seeing how lessons learned about the process of conducting joint evaluations could be applied to this experience.

CRS had planned a joint evaluation in the Yogya emergency strategy they wrote in July 2006. This decision was influenced by the ECB Project's initiative on accountability and impact measurement that had supported joint evaluations in Niger, Guatemala, and the Tsunami-affected countries during 2005 and 2006. It was also prompted by the desire of Save the Children-UK to conduct a joint evaluation.

Planning for the Yogya joint evaluation got underway in January 2007 when CRS and Save the Children jointly developed a terms of reference for the evaluation. A few months later, CARE and World Vision confirmed their interest in participating and a steering committee was formed with representatives from each of the four agencies. The agencies were driven by a motivation to learn from their response to the earthquake. All were new to Yogya, had many similarities in the types of programs they had delivered, and were affected by the same factors, such as the need to comply with conditions set forth by the government. Clearly, there was a lot to learn about how each had responded to these factors and from this, to draw conclusions about their collective response.

As the lead agency, CRS hired the evaluation team, gathered key documents from each agency, the UN, and the Government of Indonesia on the emergency response and sent them to the evaluation team. They negotiated the schedule of activities and the budget, organized logistics, and led discussions on methods with the lead evaluator. All steering committee members jointly agreed on major decisions. The agencies shared the costs of carrying out the evaluation with support from the ECB Project.

What went well

In general, the joint evaluation process went well. There was effective communication among the agency staff involved with significant trust in place. Communication infrastructure was good – there was reliable access to telephones, e-mail, instant messenger service, and all the participating agencies were located within an hour's drive of one another making face-to-face meetings relatively easy.

The lead agency played its vital management role well. The steering committee chair was successful in securing the commitment and trust of his colleagues and CRS staff did a good job in organizing all evaluation logistics, hosting the evaluation team and providing overall guidance to the evaluation team on the context and humanitarian response to the earthquake, and in the applicability of methods and questions to explore in the field.

Each of the participating agencies had monitoring and evaluation capacity, with three out of the four having monitoring and evaluation officers within their Yogya Emergency Response Teams. These staff

helped create openness within their organization to this evaluation, ensured rapid sharing of relevant documents and provided good advice to the evaluation team on methods.

The agencies also benefited from a super structure that is supportive of joint evaluations and collaboration in general. In particular, agency staff in CRS at Jakarta and headquarters level provided strong encouragement and support to staff in Yogya to lead the joint evaluation process. The culture of collaboration between these agencies promoted by ECB in Indonesia on disaster risk reduction encouraged them to try a joint evaluation.

There was a significant amount of learning and relationship building among the agencies involved in the process. The sharing of documentation and discussion when preparing for the evaluation provided an opportunity for steering committee members and monitoring and evaluation staff to learn about each other's programs and approaches as did communications during the evaluation.

Finally, the agencies made commendable and successful efforts to share the results, hosting an inter-agency event to discuss the findings of the report with local and international NGOs, local government, UN agencies, the media and universities. The report was translated into Bahasa to ensure more widespread use within Indonesia. The steering committee made plans to add graphic design elements to the report and put the main findings within a more scan-friendly PowerPoint presentation. With the intent that the report benefit the broader humanitarian community, the agencies made plans to share the report with the humanitarian learning network, ALNAP, as well as with an international distribution list.

What could have been improved

But the process was not without its shortcomings. With an overall evaluation period of 20 days, the evaluation team did not have enough time to visit enough locations for each of the four agencies. This scarcity of time was further exacerbated by the size of the evaluation team. Because the agencies wanted an independent team that would be seen as objective, they chose not to appoint representatives from each agency to the evaluation team as is usually done. In addition to the independent lead evaluator, only one staff member was assigned to the team. Instead, local facilitators, note takers and translators were hired. While excellent in their roles, local team members were relatively new to NGO work. The limited amount of emergency program experience on the team meant that specific sector areas of work were not assessed in depth.

As in-country agency staff were not assigned to be part of the evaluation team, this also limited the level of interagency learning between the participating agencies and the depth of sectoral analysis. In addition, a national consultant was not hired. This meant that the evaluation team had to depend heavily on the lead agency for advice on methods and the larger context.

Another shortcoming was related to the time of the evaluation. Coming a year after the start of the response, it was such that affected people had begun to forget what was done and by whom, what went well and what didn't.⁹

Other INGOs said they wanted to be a part of such joint evaluations and the four agencies agreed that other INGOs should have been involved, though some were initially invited and declined. As most are doing similar types of emergency response programs, there could have been great benefit to doing an evaluation of the work of all INGOs, though it would have taken more time to plan and the objectives would, perhaps, have been different.

⁹ At the same time, evaluations conducted a good while after the initial response, as this one was, can provide the opportunity to look at longer-term outcomes.

Conclusion

The agencies expect that the findings from this evaluation, being more holistic than an individual evaluation, will make a useful contribution to the humanitarian community's understanding of emergency work in Indonesia and beyond, and demonstrate their accountability to one another and the communities with whom they work.

References and Further Reading

Joint Evaluations

Guidance for Managing Joint Evaluations. DAC Evaluation Series, OECD 2006.

<http://www.oecd.org/dataoecd/29/28/37512030.pdf>

Joint Evaluations: Recent Experiences, Lessons Learned, and Options for the Future. DAC Evaluation Network Working Paper, OECD, 2005.

Lessons About Multi-Agency Evaluations: Asian Tsunami Evaluation Coalition. http://www.tsunami-evaluation.org/NR/rdonlyres/9DBB5423-E2EF-43AB-B6D2-2F5237342949/0/tec_lessonslearned_ver2_march06_final.pdf

General Evaluations

USAID Center for Development Information and Evaluation, Performance Monitoring and Evaluation TIPS series.

<http://evalweb.usaid.gov/resources/tipsseries.cfm>

Western Michigan University, Evaluation Center. Evaluation Checklists

<http://www.wmich.edu/evalctr/checklists/checklistmenu.htm#models>

Shoestring Evaluation: Designing Impact Evaluations Under Budget, Time, and Data Constraints. M. Bamberger, J. Rugh, M. Church, and L. Fort, *The American Journal of Evaluation*, 2004.

Utilization-Focused Evaluation Checklist. Michael Quinn Patton

<http://www.wmich.edu/evalctr/checklists/ufe.pdf>

¹ Joint Independent Evaluation of the Humanitarian Response of CARE, CRS, Save the Children and World Vision to the 2005 Food Crisis in the Republic of Niger, November 2005.