Real-Time Evaluation in Humanitarian Emergencies

Emery Brusset, John Cosgrave, Wayne MacDonald

Abstract

The authors describe real-time evaluation (RTE) as a specific tool in disaster management and within the literature on formative evaluation, monitoring, and impact assessment. RTE offers the possibility of exploring innovative ways to empower frontline disaster response staff, possibly even beneficiaries of assistance. The authors describe conditions for the success of RTE, including field credibility, organization, and rapid analysis. © Wiley Periodicals, Inc., and the American Evaluation Association.

Real-time evaluation (RTE) is an approach increasingly invoked to generate knowledge in international humanitarian emergencies. It is defined as “an evaluation in which the primary objective is to provide feedback in a participatory way in real time (i.e., during the evaluation fieldwork) to those executing and managing the humanitarian response” (Cosgrave, Ramalingam, & Beck, 2009, p. 10). In particular, it provides a fresh view to managers and concentrates to a greater extent on the immediate effects of the intervention than normal monitoring would.

Interagency or joint RTE is also increasingly used by agencies operating in disasters, and the United Nations has commissioned a series of such
evaluations (Cosgrave, Gonçalves, Martyris, Polastro, & Sikumba-Dils, 2007; Inter-Agency Standing Committee [IASC], 2006; Turner, Baker, Oo, & Aye, 2008; Young, Khattak, Bengali, & Elmi, 2007). Joint evaluation is perceived as particularly apt to study issues that cannot be studied usefully by any single agency evaluation (e.g., coordination, joint programming, gaps and balance between sectors). RTE allows for the assessment of a joint strategy even at the formative stage.

Why Are Real-Time Evaluations Necessary?

RTE refers to methods that emphasize light and mobile evaluation teams. It is focused on interactive local learning and on rapid feedback to stakeholders (either on the ground or more remote) for purposes of learning as well as of accountability. RTEs are uniquely positioned to answer the formative questions of interventions: What has been the performance so far and what is the optimal use of resources? What direction should things take? How should information be succinctly presented to mobilize resources? What stages have been reached and what should the next ones be? When and how should we begin to withdraw or begin another intervention?

Some of these questions deal not just with a valuing of what has happened, but also to an estimation of the likely outcomes of the current course of action, what the U.S. General Accounting Office (GAO) refers to as “prospective evaluation” (GAO & Datta, 1990). However, there remains no clearly defined position within the international evaluation community and among humanitarian agencies regarding the nature of RTE and the implications of key choices that RTE requires if it is to achieve its potential.

Contrasts With Other Evaluation Approaches. A review of the World Food Program’s (WFP) experience with RTE notes that the complex environment and the different perceptions within the agency of what an RTE is makes RTEs arduous (World Food Program, 2004). What are the risks? To answer this question we first highlight ways in which RTE differs from other forms of performance assessments.

RTE is different from audit and inspection, mainly because of the breadth of the subject and the absence of universal standards. Both audit and inspection can be described as more or less detailed checks on the compliance of an operation to a norm that is verifiably stated in organizational, industry, or sector standards, with a focus on management structures, finances, and procurement. The aim of inspection and audit is risk reduction, and the primary purpose is accountability. RTE, on the other hand, examines the value of a result (and is hence more oriented to external standards of performance as the point of reference, even when these standards are not fully explicated) and explores the reason for achieving or not achieving those standards with a view to improvement.

RTE is distinct from other methods such as After-Action Reviews (a term coined by the U.S. Department of Defense) or Program Performance
Reviews (U.S. Office for Transition Initiatives) because of its emphasis on a broad learning perspective in the midst of a response and the systematic use of criteria-based analysis. After-Action Reviews are post hoc structured reviews or debriefings by participants and those responsible for the project or event for analyzing what happened, why, and how it can be done better. Both RTE and After-Action Review use supportive peer presence in the form of experienced facilitators cum reviewers, but After-Action Review takes place in an environment removed from the actual operation and RTE is simultaneous with program implementation. It is possible that RTEs could use After-Action Reviews as data-gathering methods.

Monitoring is usually defined in humanitarian assistance as the periodic processing of information on outputs and outcomes along a preestablished plan, most usually focused on key performance indicators. This function is performed by personnel in country or close to the operation, and supports project reporting. RTE is, on the other hand, an occasional tool, carried out by persons external to the operation. Even though they may be agency employees, these individuals are not usually involved in the implementation structure for reasons of impartiality and use of time.

Needs assessment is obviously related to how well the population and the state are doing, and RTE is a way of assisting in that process by complementing findings. Needs assessment is often a very technical task, which can benefit from inputs provided by personnel with a more systemwide performance focus. Evaluation done in real time is a subset of more general standard humanitarian or emergency evaluation—with a particular twist, which is to capture humanitarian action as it is occurring. Guidelines on the evaluation of humanitarian action (Beck, 2006; Hallam, 1998; Organisation for Economic Co-Operation and Development, Development Assistance Committee [OECD/DAC], 1999) or evaluations conducted in conflict environments (OECD/DAC, 2008) point to the importance of the specifics of a situation (e.g., polarization of society, difficulty finding information, poorly documented planning, changes in circumstances) that require specific accommodations of methodology. RTE is confronted with the same conditions, albeit accentuated, in a short reporting time.

The more theoretical literature published to date on this subject indeed would not claim that RTE is a fundamentally new concept, but rather a codification of a particular form of formative and process evaluation. For example, RTE is defined in the OECD/DAC glossary (2002) as: “An evaluation intended to improve performance, most often conducted during the implementation phase of projects or programs. This is contrasted with process evaluation, which is an evaluation of the internal dynamics of implementing organizations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these.”

Scriven (1991) defines formative evaluation as typically conducted more than once, for in-house staff, with the intent to improve. In an informal sense, operational personnel are constantly doing formative evaluation, but
this is not done with the key additional assets of independent views and systematic analysis of performance information. The formal nature of RTE, as a type of formative evaluation, stems from the accountability function of reporting back to senior management and public stakeholders on the relationship between current and optimal use of resources.

**A Particular Relation to Accountability and to Learning.** Learning and accountability are concepts that drive the specific choices regarding the conduct and use of evaluation. The international evaluation community and emergency organizations often see these concepts as inextricable, mingled into the very make-up of the exercise. However, learning and accountability are distinct.

With respect to learning, RTE is a useful tool to improve management decision making at all levels. The recommendations and dynamics of RTEs (for example, of discussion and conversation carried out by the evaluators, but also by stakeholders independently of the evaluators) are generally seen as more relevant for operations than a longer-term evaluation.

Accountability is also a high-profile aspect of evaluation in emergency and disaster management, mostly because of the public dissemination of reports, but also because evaluators are frequently seen as channels of information to headquarters and donors. Accountability is sometimes a constraining factor, partly for reasons of sensitivity and hence access to information (an issue to which we will return), but also because of the care that should go into developing and presenting evaluation findings. Some emphasize the need for independence of RTE (in essence those who evaluate not being the same as those who implement, and being more impartial). Others emphasize that the quest for independence and accountability should not be an obstacle to fast execution of RTEs, and that truly external review will necessarily slow down the process.

The learning function of the RTE occurs through influence on policy makers, but also has clear value for those on the ground. In these conditions the United Nations, which fielded an RTE of the Darfur operations (Broughton et al., 2006), explained that evaluation is likely to be seen by country teams as a supportive measure that can help to adjust planning and performance.

These points highlight two underlying dilemmas in RTEs. One is whether RTE should be aimed at learning at the immediate level of operations, and downward or lateral accountability (and should be decided and timed by the managers), or whether they should be triggered by bodies external to the operation for broader corporate decision making, technical programming, and accountability purposes. This has implications for the scope of work and methodology, which will be reviewed in more detail in the next section.

The second dilemma is the choice between single-agency RTEs, which present fewer constraints in the area of accountability, and joint or interagency RTEs, which provide opportunities for broader perspectives and a deeper
understanding of cross-cutting needs and response regime. This implies further considerations for the steering and purpose of the RTEs, as well as validation and communication strategies. The following sections tackle these issues from different perspectives.

**Methodological Issues of Real-Time Evaluations**

RTE is a subset of humanitarian evaluation, which is itself a subset of development evaluation. Although the real-time format imposes some constraints, the nature of humanitarian emergencies, their normal locations, and the nature of humanitarian response also limit methodological choices. Most large-scale humanitarian emergencies take place in developing countries. The conditions in developing countries limit the use of statistical sampling from a known sample frame, as there are often no reliable sample frames of the general population. The low level of telephone coverage and the absence of standard addresses rule out many surveys.

Another constraint is the nature of international humanitarian emergencies themselves. By definition, these are sudden events and responses are essentially unplanned in strategic terms. Each element of the humanitarian community acts independently while attempting to coordinate its interventions with other actors. The lack of coordination is often highlighted in the evaluations of major responses.

Furthermore, baseline studies are usually nonexistent, making before and after comparisons impossible. Even when organizations develop plans, they repeatedly have to change them as the responding organization learns more about the context of the affected community, or what was planned becomes redundant due to actions by the government or other organizations. Thus, it is very difficult to compare actual activities against planned activities in humanitarian response, as the planned activities may change even on a daily basis.

Not only does this make impact very difficult to measure, but the chaotic jumble of international organizations makes it very difficult to attribute any observed impact to any cause. If broader aims, such as preventing outbreaks of communicable disease, are met, what are the reasons? Is this because of medical coverage, the supply of clean water, effective epidemiological surveillance, or vector control? Or, is it because such outbreaks are extremely rare in any case (Floret, Viel, Mauny, Hoen, & Piarroux, 2006)? If an impact is attributed to a particular intervention, then which of the organizations working on that intervention is most responsible?

The aforementioned issues mean that humanitarian evaluation largely relies on a range of (predominantly qualitative) methods including key informant interviews, document research, discussions with communities, and occasional surveys. Although surveys are used in some specific areas of humanitarian response such as nutrition, their use in general evaluation...
of humanitarian response is fairly limited. Notably, the use of this method increased after the December 2004 Indian Ocean tsunami response. (See Spence & Lachlan, this issue, for a discussion of strategies for using survey methods in the evaluation of disaster relief.)

RTEs in the early stage of humanitarian response, that is, when they have the greatest opportunity to influence the overall humanitarian response, bring additional constraints. Four constraints specific to real-time humanitarian evaluation in the early stages of an international response are as follows:

1. The heavy workload on humanitarian staff dealing with the response leaves them little time for the evaluation team.
2. There is a lack of analytical documentation available at the early stage.
3. The field work needs to begin quickly to be most useful.
4. The time for analysis is limited.

The time limit for the data analysis arises as the usual intent to provide a draft report before leaving the field or at least a thorough briefing for the team implementing the humanitarian response.

Surveys in the early stages are very difficult because of the lead time needed to develop, translate, and test survey instruments, and because of the time needed to train enumerators. This preparation time effectively rules them out from use in early-stage RTE. Also, document review tends to be limited by the descriptive nature of documents. Although a later humanitarian evaluation may be able to utilize a wide range of documents in the analysis, documents in the initial stages tend to be descriptions of the ongoing response and of the intent of the responding agency. However, even at the early stages there may be a very large number of available documents. These can be a rich source for some types of analysis (including the presence of particular terms or key words in the documents) or the balance between the volume of documents from different types of actors.

Some of the questions that may be asked of an RTE relate to making assessments on likely outcome of current policies, a form of prospective evaluation. The General Accounting Office (1990) provides a useful model for evaluators making such an assessment. This model, presented in Figure 2.1, depicts the relationship between conceptual, empirical, and operational elements of a prospective evaluation.

Evaluators can only make a prospective evaluation of current policies in an RTE if they have knowledge not only of the underlying concepts, but also of the operational constraints within the agency, and of the history of such interventions. Thus, although evaluators are sometimes asked to make prospective judgments in RTEs, they should refrain from doing so unless they are confident in understanding all aspects of this triad.

**Specifics of Data Collection.** As with humanitarian evaluation generally, the main method used for the collection of information in humanitarian
RTE is the semistructured key-informant interview that is analyzed in terms of salient themes. In the humanitarian context, such interviews are conducted with key informants purposively selected because they are “information-rich” individuals (Patton, 2002, p. 40). Evaluation teams may supplement their initial, purposive sampling with snowball sampling where each interviewee is asked to recommend other key informants until the accessible pool or the evaluators’ time is exhausted. Snowball sampling is particularly applicable for RTEs where there is often no single contact list available.

Evaluators typically use an interview topic list for key informant interviews, which last from 45 to 60 minutes. They may work individually or in pairs to conduct the interview, which may be with individuals or small teams, and normally take notes that may be transcribed later. One constraint for humanitarian RTE is that interviewees may have to respond to urgent operational issues that limit their availability or lead to the premature ending of interviews. To cope with this, real-time evaluators need to be flexible in their scheduling and their approaches. For example, key questions may have to be asked at the beginning of an interview rather than waiting until the rapport is established. After-Action Review has been used for some humanitarian reviews (Baker, 2005). An After-Action Review is a facilitated discussion of an event that leads the participants through an analysis of what happened, why it happened, and how to sustain strengths and improve on weaknesses.

The workload on international field staff means that the evaluators must have field credibility in terms of their experience with humanitarian response. Such experience is also essential for the evaluators to make the most of the opportunity to observe the actions undertaken as part of the humanitarian response.

Interviews with beneficiary groups are an important source of information. In many development settings, people have to make a contribution for the goods or services provided by the organization. Their willingness to

Figure 2.1. The Triad of Analysis for Prospective Evaluation

![Diagram of the Triad of Analysis for Prospective Evaluation]

- **Empirical**: Historically, has it worked?
- **Conceptual**: Logically, should it work?
- **Operational**: Practically, could it work?
do so provides feedback to the organization on whether the goods and services are relevant. If they are not relevant, people will not make any contribution. By contrast, in humanitarian responses, goods and services are usually provided without any beneficiary contribution, and people will accept things that they don’t need because they are free. This means that agencies may not get feedback on how useful their assistance is. This is especially true in the early stages of a humanitarian response and makes group interviews even more useful for RTE.

Group interviews usually take the form of meetings at which evaluators pose questions to a collection of anywhere from a handful to hundreds of community members. The presence of other community members acts as a social control on those responding to the questions, and others will often add their opinion if they disagree with what someone else has said. These group interviews are very distinct from focus-group discussions as defined by Krueger and Casey (2009). Focus groups require more preparation and control than can usually be achieved in RTE.

Observation is another key method in RTE. It is very valuable when the evaluation happens during the early stages of a response, as operations are at full strength and issues may be clearer than at other times. It is also important because operational staff may not have had time to analyze their own observations and internalize lessons that they can articulate during key informant interviews. Observation is a key resource for triangulation, and triangulation is the chief means by which RTEs strive for accuracy and reliability. Triangulation seeks to compare data gathered by different paths to determine how accurate and reliable a picture the data present of the situation. Triangulation can compare data from different sources, methods, areas, or evaluators.

The need for an RTE team to provide feedback before leaving the field is a significant constraint on the depth of analysis in which the team can engage. In an international humanitarian evaluation generally, the writing of a report after the fieldwork allows time for reflection and for considering the evidence to develop a deeper analysis, sometimes supported by further documentary research on specific aspects. In contrast, an RTE team must conduct its analysis before leaving the field.

One way of approaching this problem is to use a tool to organize the evidence uncovered during the evaluation to facilitate the drawing out of findings, conclusions, and recommendations. (See Table 2.1.) By explicitly linking evidence to findings, findings to conclusions, and conclusions to recommendations, such a tool can help to ensure that there is a strong chain of evidence for the evaluation’s eventual recommendations.

Building a strong chain of evidence in this way increases the reliability of the information presented in the evaluation, a point made by Yin (2003) concerning case studies generally. Ideally, every evaluation finding should be supported by several different types of information, and every conclusion supported by several findings. Using such a tool also allows the evaluators to see when they have reached saturation on a particular issue. This happens
<table>
<thead>
<tr>
<th>Evaluation Question/Issue</th>
<th>Evidence About This Issue</th>
<th>Finding</th>
<th>Conclusion</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation question or issue emerging from the fieldwork</td>
<td>A concrete statement from interview, documentation, or an observation</td>
<td>Based on several pieces of evidence</td>
<td>Based on several findings</td>
<td>Based on one or more conclusions</td>
</tr>
<tr>
<td>Evaluation question</td>
<td>From interview 1</td>
<td>From site visit 2</td>
<td>From interview 3</td>
<td>From document 5</td>
</tr>
<tr>
<td>Evaluation issue (an issue emerging from fieldwork)</td>
<td>From interview 9</td>
<td>From interview 11</td>
<td>From document 7</td>
<td>From document 15</td>
</tr>
</tbody>
</table>
when new evidence on any issue continues merely to replicate previous evidence rather than to add any new depth to the issue. Identifying when you are reaching saturation on particular issues is useful for an RTE team, as it allows the team to use its time effectively to concentrate on a range of critical issues.

**The Promise of Real-Time Evaluation**

RTE can be seen as a practice that is both promising and increasingly popular in the international humanitarian field. However, there remains a need for considerable reflection regarding who it is for, who will control it, and how it should be carried out. The answers will depend on the situation. As Sandison (2003) noted in her early review of the process of RTE, the main value of RTEs is through “... its timing and rapid feedback, and its combination of proximity to the actual programme and people with its evaluative distance and objectivity” (p. 8). This combination of an external view with the ability to influence decisions as they are made is the real attraction of RTE.

The days of formative evaluation as something limited to only one organization—a preoccupation of in-house program/project staff—may be numbered. In differentiating formative from summative evaluation, Scriven used a simple analogy when he quoted Robert Stake: “When the cook tastes the soup, that’s formative; when the guests taste the soup, that’s summative” (1991, p. 169). RTE blurs this distinction. Scriven (1991) cautions evaluators to avoid “loading formative evaluation with obligations that will bog it down, or expectations that will be disappointed” (p. 169). However, RTE does not entirely embrace this advice. In fact, the cook in the earlier analogy now starts to look more like a team of sous chefs, often involving the guests in the kitchen (e.g., senior management, decision makers, communications experts, external audiences, members of the affected population) to taste the soup. If not to their palate, ingredients can be changed or possibly the entire effort discarded.

As expectations mount, the RTE menu becomes more complicated. This may be reflective of the enlarged number of actors around the stove. Whatever the reason, the task of coordinating and using internal and external RTE evaluation teams becomes more challenging. Timing and rapid feedback are important considerations. Indeed, disaster management systems are expected, and even obliged, to demonstrate credible results quickly on the relevance and effectiveness of actions taken. Within this context, a focus on value for money, linking timely disaster spending to priorities and achievement of tangible outcomes, is no longer subject to negotiations. It is a given and part of the very fabric of a renewed interest and preoccupation with good governance.

RTE can be more complicated, and sometimes more resource intensive, than other evaluations, but it can provide a unique and opportune window...
on critical system level concerns, important indirect effects, and other disaster management issues. Above all, it provides a systematic, evidence-based analysis to counteract the hasty and erroneous perceptions that often emerge in emergency response.

References


EMERY BRUSSET is director of Channel Research, a research consultancy firm specializing in evaluations.

JOHN COSGRAVE is a professional evaluator and trainer in the humanitarian sector with InterWorks Europe.

WAYNE MACDONALD is director of Corporate Performance and Evaluation with Canada's Social Sciences and Humanities Research Council in Ottawa.