Evaluating Humanitarian Action in Real Time: Recent Practices, Challenges, and Innovations

Riccardo Polastro

IOD Parc

Abstract: *Real-time evaluations (RTEs) are formative, utilization-focused evaluations that provide immediate feedback. Within the humanitarian system, RTEs are an innovation for improved learning and accountability today. In fluid and fast-changing environments they bridge the gap between conventional monitoring and evaluation, and influence policy and operational decision-making in a timely fashion. RTEs identify and propose solutions to operational and organizational problems in the midst of major humanitarian responses. The article (a) defines RTEs and situates them in the wider evaluation landscape; (b) examines RTEs' use and users; (c) focuses on current methodological approaches; (d) looks into challenges, opportunities, and limitations that condition uptake; and (e) draws lessons and recommendations.*

Keywords: *learning and accountability system-wide evaluations, real-time evaluation, utilization-focused*

Résumé : *Les évaluations en temps réel (ÉTR) sont des 'évaluations formatives axées sur l'utilisation qui donnent de la rétroaction immédiate. Aujourd'hui les ÉTR sont une innovation qui contribue à améliorer l'apprentissage et la redevabilité dans le système humanitaire. Dans des contextes fluides qui changent rapidement, elles surmontent les lacunes entre le suivi et l'évaluation conventionnels et influencent les politiques et les prises de décisions opérationnelles de façon opportune. Les ÉTR identifient et proposent des solutions aux problèmes opérationnels et organisationnels au cours de réponses humanitaires majeures. Cet article (a) définit les ÉTR et les situe dans le champ plus étendu de l'évaluation; (b) examine l'utilisation et les utilisateurs des ÉTR; (c) analyse les approches méthologiques actuelles; (d) étudie les défis, les possibilités, et les limites qui affectent la mise en œuvre; et (e) présente des leçons et des recommandations.*

Mots clés : *évaluations de l'apprentissage et de l'obligation de rendre compte à l'échelle du système, évaluation en temps réel, évaluation axée sur l'utilisation*

Corresponding author: Riccardo Polastro, riccardopolastro@gmail.com

INTRODUCTION

Contextualizing Real-Time Evaluations

Multiple definitions of humanitarian action exist, but for the purpose of this article we will use the one provided by the Good Humanitarian Donorship Initiative (2003). Humanitarian action is the action taken to "save lives, alleviate suffering and maintain human dignity during and in the aftermath of man-made crises and natural disasters, as well as to prevent and strengthen preparedness for the occurrence of such situations."

When the Cold War ended, the number of agencies protecting and assisting people in countries affected by conflict and/or disasters could often be counted on one's hands. Funding resources for humanitarian action were limited and their proportion of official development assistance (ODA) was relatively small. Evaluation[1] of humanitarian action was sporadic.

Today, the situation is flipped and the picture is polarized:

- Humanitarian action resources allocated on a yearly basis have grown up to US$17 billion (Poole et al., 2012). For some OECD countries that are members of its Development Assistance Committee (DAC), the proportion of humanitarian action funding represents 20% of their ODA.
- There are countless actors involved in responding to crisis, and the number tends to mushroom after each major crisis, as observed in the Great Lakes genocides, the Former Yugoslavia fragmentation, the 2004 tsunami in the Indian Ocean, the 2010 Haiti earthquake, the 2013 Philippines Haiyan typhoon, or in the last three years in the countries bordering Syria.
- Every year millions of people, often in the world's poorest countries, are assisted and protected to survive and recover from conflicts and disasters.
- From being exceptional, humanitarian action evaluation is a "booming industry."

According to the ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action) evaluation of humanitarian action pilot guide, the evaluation of humanitarian action can be defined as "the systematic and objective of humanitarian action, intended to draw lessons to improve and practice and enhance accountability" (Buchanan-Smith & Cosgrave, 2013). However, humanitarian action has historically been subjected to less rigorous and extensive evaluation procedures than development aid, and has been plagued by poor monitoring and "methodological anarchy" (Hallam, 1998; OECD DAC, 1999, p. 2).

In humanitarian settings objectives are fast-changing, as the situations on the ground are fluid and rapidly evolve (e.g., continuous civilian population displacements, epidemic outbreaks), planning tends to be poorly documented, and baseline data are often absent. When conflicts prevail, security and access

tend to be highly deteriorated. As a result, despite multiple efforts, performance and learning can stall and poor accountability to national stakeholders and aid recipients prevails.

Following years of meta-evaluation efforts that reviewed humanitarian evaluations, the limitations of longer and more costly ex-post evaluation efforts soon became evident. Evaluation professionals asked themselves if "learning could be achieved in a mo re timely and economical fashion." During ALNAP[2] biannual meetings, real-time evaluation (RTE) working groups were established, where evaluation professionals from the International Red Cross Movement, UN agencies, international nongovernmental organizations (INGOs), independent consultants, and donors exchanged views and approaches putting momentum behind RTEs and breaking new ground in humanitarian evaluation.

The urgent nature of humanitarian response rapidly boosted RTE's uptake by key humanitarian agencies that, until then, had co nsidered that traditional ex-post evaluations came too late to influence the operations while they were still ongoing. As RTEs immediately provided suggestions for improvement, new humanitarian evaluation architecture progressively emerged, where RTEs became a central pillar.

During the first Gulf War, following the massive displacement in Northern Iraq, an RTE prototype was trialed by the United Nations High Commissioner for Refugees (UNHCR) in 1992; ho wever, it was only after the Kosovo War (1998–1999) that RTEs were formally introduced and planned for. Since then, key humanitarian actors tested RTEs and piloted their approach. In recent years, the number of RTEs has exponentially grown, and on average 12 RTEs have been fielded each year since 2010. Nowadays RTEs are either commissioned individually or jointly by donors, UN agencies, the International Red Cross Movement, and INGOs. RTEs are carried out by either independent or internal evaluation teams.

Today, a single-agency RTE may cost between US$12,000 and $70,000, while a joint RTE may cost between US$70,000 and $150,000. What increases costs are the length of evaluation; the geographical scope of the evaluation; the size of the team (generally from one to four consultants) and the number of internal consultants as well as international and national consultants; the number of pre- and post-evaluation missions done by the commissioning agency; and the number of field visits and debriefings. Nevertheless, the cost of RTEs, even when they are at their highest, is proportionally small compared to the cost of the humanitarian responses they look into, and they potentially can contribute to improving the efficiency and effectiveness of the overall humanitarian system response to a crisis (Telford, 2009).

WHAT ARE RTEs?

When looking at the evaluand, one understands that RTEs aren't a new model or theory of contemporary evaluation, but rather a new categorization of a particular form of formative evaluation (Scriven, 1991) and process evaluation. RTEs are

formative evaluations of intermediary results. Patton (2008) classifies RTE as a development evaluation with the main function of improving rather than proving, that can be applied when a program is in a continuous state of change. RTEs are to be considered as one option in a utilization-focused[3] process.

For the purpose of this article we will look at RTEs as "a subset of humanitarian evaluation, which is itself a subset of development evaluation" (Brusset, Cosgrave, & MacDonald, 2010, p. 13). However, humanitarian evaluation as recognized tends not to be well informed by mainstream conceptual frameworks nor has it developed its own yet (Feinstein & Beck, 2006). RTEs have been commonly practiced for years in various fields of science and technology ranging from computing science to health and climate change.[4]

When looking through the evaluation spectrum, it is easy to distinguish RTEs from monitoring, midterm, and ex-post evaluations (see Figure 1).

Monitoring is often absent in humanitarian settings and, when it is in place, is not always adapted to the changing realities on the ground. An RTE can help bridge the gap, as it provides an immediate snapshot that can help managers identify and address the strengths and weaknesses of the response. Furthermore, RTEs can also assess how well the monitoring systems are working.
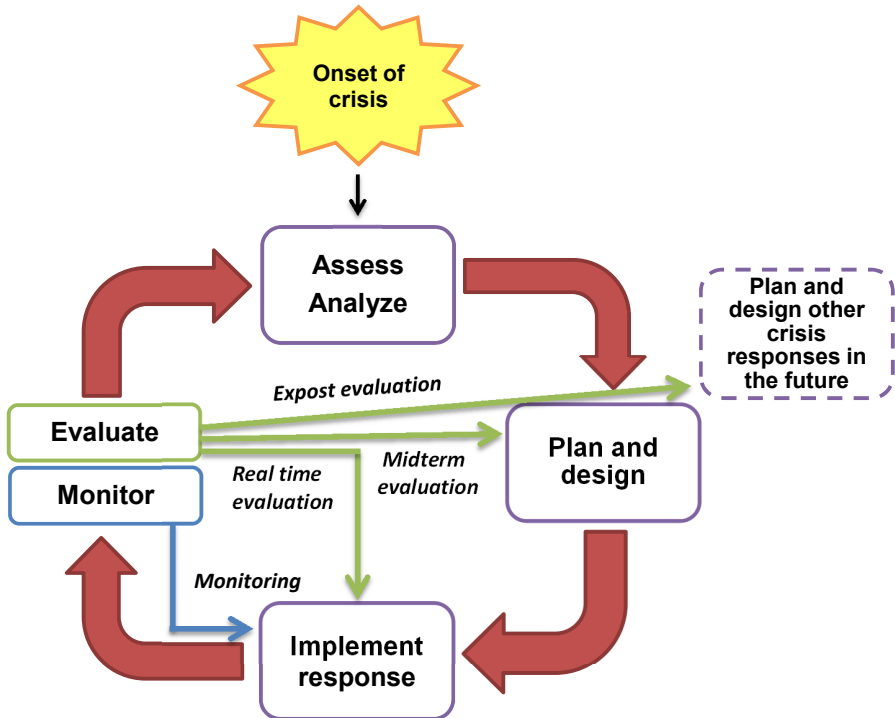

Figure 1. Evaluation and monitoring in sudden-onset crises.

However, RTEs are different from monitoring, in that they

- are prospective and look at the likely outcomes of current policies, not simply keeping track of whether targets are being met,
- are intermittent and episodic, while monitoring is continuous and routinely gathered,
- use and externally define evaluation criteria, while monitoring generally applies the project management framework and predefined performance indicators to measure the extent of progress and achievement of objectives of the specific project/program/response,[5]
- are carried out by external and independent consultants, while monitoring is generally carried out by the organization's staff, and
- are used by the internal management team, donors, and other stakeholders, while monitoring is used by the internal management.

Another distinguishing feature of RTEs is that they are conducted during implementation rather than responding after programs are closed. RTEs look at "today" to influence "this week's" programming and act as immediate catalysts. Therefore, they must be rapid, flexible, and responsive. In contrast, midterm evaluations look at the first phase of the response to influence programming (planning and design) in its second phase, while lessons from ex-post evaluations can be used in future programs when responding to crises (Cosgrave, Ramalingam, & Beck, 2009).

When Are They Carried Out and What Are They For?

As mentioned above, the defining characteristic of an RTE vis-à-vis a standard evaluation is timing: RTEs are, in principle, carried out in the midst of an emergency operation. When fielded too late they become more similar to midterm evaluations. RTEs "encompass strategic, process and performance approaches to evaluation, rather than impact evaluation, simply by virtue of its timing" (Sandison, 2003, p. 7).

They can free up operational bottlenecks and provide real-time learning. RTEs are also improvement-oriented reviews—dynamic tools used to adjust and improve planning and performance as the disaster response unfolds. When recurrently and systematically done, they can contribute to reinforcing accountability to beneficiaries, implementing partners, and donors, and can bridge the gap between monitoring and ex-post evaluation.

An RTE is intended to be a support measure for learning in action. They are interactive, involving a wide range of stakeholders and therefore contributing to peer-to-peer learning and accountability.[6] Because the results and recommendations are intended to be applied immediately, RTEs must be rapid, flexible, and responsive.

RTEs can also reinforce the link between operations and policy formulation. Following the 2007 Mozambique floods (Cosgrave, Gonçalvez, Martyrs, Polastro, &

Sikumba-Dils, 2007), an RTE examined how the UN humanitarian reforms were being rolled out in the field. A management matrix was implemented and the recommendations were closely monitored by the UN Emergency Relief Coordinator, looking at how Humanitarian Country Teams were applying lessons on UN humanitarian reform (Beck and Buchanan-Smith, 2008).

Furthermore, in response to the failings (well documented by RTEs) of the international community's response to the 2010 Haiti earthquake and Pakistan floods, the heads of UN agencies formally adopted the "transformative agenda" to address the operational challenges related to the humanitarian reform process.[7] When publishing the interagency report of the humanitarian response to the 2010 Pakistan floods, Valerie Amos, the Under-Secretary General for Emergency Response, noted, "The ability to improve the humanitarian system, and ultimately to respond more effectively to future humanitarian crises, will depend on our ability to learn from real time lessons, e.g., by taking immediate action on evaluations and reviews."[8]

WHO ARE THEY FOR?

An RTE is a participatory evaluation that is intended to provide immediate feedback during fieldwork. In an RTE, stakeholders are primarily executing and managing the response at field and national levels, but also at regional and headquarters levels. In this type of evaluation the country team and the field staff have a greater stake. An RTE provides instant input to an ongoing operation and can foster policy, organizational, and operational change to increase the effectiveness and efficiency of the overall disaster response (Jamal & Crisp, 2002).

Normally, the primary audience of an RTE is managing the response in the field, the secondary audience is at headquarters, and the tertiary audience is the humanitarian system as a whole. However, the stakeholders' stakes strongly depend on who initiates the evaluation and who raises the key issues to be addressed. If the evaluation is launched from headquarters, the level of ownership in the field is likely to be reduced. In this case, the RTE may be perceived as intrusive and geared primarily to upward accountability rather than facilitating learning on the ground. In contrast, when the exercise is initiated in the field (as was the case in the Mozambique interagency RTE of the response to the floods and cyclone in 2007 [Cosgrave et al., 2007] and in the humanitarian response to Pakistan's internal displacement crisis in 2010 [Cosgrave, Polastro, & Zafar, 2010]), the RTE is usually welcomed, as all actors believe that it can contribute to improving the ongoing response and unlock operational bottlenecks.

An RTE can contribute to improved accountability to different stakeholders by involving them throughout the evaluation process. With both interagency and single-agency RTEs, the agencies whose activities are being evaluated are meant to act on the recommendations. However, feedback tends to be given mainly to peers and donors. Despite being the primary "clients" of the humanitarian aid industry, beneficiaries and local and national governments rarely receive feedback on the recommendations or how they are being implemented.

Single-agency versus Interagency RTEs

Single-agency RTEs focus on a particular agency response, while interagency or "joint" RTEs and system-wide evaluations evaluate the response of the whole humanitarian system to a particular disaster event or complex emergency. Joint RTEs adopt a broader perspective and deeper understanding of cross-cutting elements such as the overall direction, coordination, and implementation of the response, including needs assessments, threats to humanitarian space, coordination, joint programming, gaps, and operational bottlenecks. Joint RTEs tend to look at the humanitarian system from a helicopter view perspective; they look at the whole forest rather than specifically at group of trees or the leaf, as partial system evaluations and program evaluations do.

Furthermore, as noted in the OECD DAC guidance document, joint evaluations have the potential to bring benefits to all partners. Collaborative working offers opportunities for mutual capacity development and learning between the partners, for building participation and ownership, for sharing the burden of the work, for increasing the legitimacy of findings, and for reducing the overall number of evaluations and the total transaction costs for partner countries (OECD DAC, 2006).

When done jointly, an RTE represents a learning opportunity for participating agencies and national and local governments, as well as affected communities. Actors involved in the response (the affected population, national government, local authorities, the military, local NGOs, international donors, the UN, the Red Cross/Red Crescent, and INGOs) are consulted, fostering increased learning and accountability across the humanitarian system. Broad participation increases ownership of findings and makes follow-up on recommendations more likely.

SOME METHODOLOGICAL APPROACHES

Evaluating humanitarian responses provided in uncertain, turbulent, fluid, and insecure environments presents challenges beyond those encountered under more stable conditions. This is mainly due to issues of access and security, and the frequent absence of standardized monitoring and comparable data sets on the response to the affected population (Polastro, Khalif, et al., 2011).

Data are very sensitive, scattered, and atomized. In situations of conflict the reliable information—notably statistics—that is available is often of insufficient quality to make any reliable assessment.[9] To tackle this, humanitarian evaluation teams use various data collection methods through an inclusive and participatory process, attempting to get as many stakeholders as possible involved in the evaluation.

As with other humanitarian evaluations, rapid methods are necessary for formative evaluation purposes. RTEs essentially use qualitative methods, including stakeholder analysis, extensive interviews both with aid providers and aid recipients (snowball sampling[10] with "information-rich" individuals, group discussions, etc.), broad field travel to sample sites, wide-ranging observation and

documentary research, thematic analysis, evidence tables to collect systematically the data, and workshops to validate findings and prioritize recommendations.

As in other evaluation practice, triangulation is key to ensure data validity and robustness of the evidence collected. A matrix of evidence is a quick, useful, and systematic way to triangulate and cross-validate findings. Before drawing up the findings, teams should cross-validate the information:

- documents against interviews;
- research/documentary evidence against interviews;
- observation against interviews;
- comments against initial findings presented during workshops

According to the Inter-Agency RTE "lessons learnt" workshop that took place in Geneva in November 2010, an evidence-based RTE can serve as a tool for change. As noted in other research, the credibility of evidence is linked to the credibility of its source (e.g., a respected expert or organization) and to its fit with professional and practice wisdom (Nutley, Walter, & Davies, 2007).

Nevertheless, inconsistencies in methodology, an absence of theoretical underpinnings, and resistance to categorization and standardization have left some in the evaluation community wondering whether RTEs are sufficiently rigorous tools (Herson & Mitchell, 2005). To tackle this, major humanitarian agencies such as UNHCR, UNICEF, and the World Food Program, as well as ALNAP and the Office for Coordination of Humanitarian Affairs (OCHA) have developed specific concept papers, lessons learned, and methodological guidance. The overall quality of an evaluation depends on six key factors: design (purpose and approach), participation and ownership, planning, evidence, follow-up mechanisms, and evaluator credibility (Sandison, 2006).

One key success factor is using methods to create ownership and foster learning and participation among the agencies involved in the response. Studies suggest that personal contact and face-to-face interactions are the most likely to encourage policy and practice use of evidence, which suggests that evidence use may above all be a social process (Nutley et al., 2007).

According to Bamberger (2008), one of the key determinants of whether an evaluation will be used is the extent to which its clients and stakeholders are involved in all stages of the evaluation process. To be able to make evaluation evidence usable, the interests of the stakeholders must have already been correctly identified at the beginning of the evaluation (Carlsson, 2000). Practice suggests that getting policy-makers in headquarters to own and use evidence involves getting commitment and buy-in at the most appropriate levels.

From the outset, the methodology should be designed to ensure adequate and regular feedback to country-level staff. When considering methods, the team should choose ones that are adapted to the context and can be understood by the country team.

Furthermore, many potentially useful evaluations have little impact because the findings are not communicated to potential users in a us eful or comprehensible way (Bamberger, 2008).When doing an RTE, the team should prioritize facilitated, reflective, participatory methods to the point of developing and delivering final recommendations with the main stakeholders in the affected countries. Dissemination strategies must be tailored to the target audiences, reaching them with timely, relevant information backed up by sound evidence.

Today evaluation managers of different organizations commissioning RTEs are progressively using more evaluation briefs or short summaries, translations of the executive summary into the local language, video clips,[11] and other tailored communication tools to reach different audiences with relevant findings.

To be utilization-focused, the evaluation team needs to understand, before travelling to the field, what evidence needs to be collected and for whom. An RTE is more interactive than other types of evaluations—the evaluator acts as a facilitator and must engage in a sustained dialogue with key stakeholders throughout the evaluation at field, capital, regional, and HQ levels. The level of interactivity must be high and continuous to identify and resolve problems with organizational or operational performance and to act as a catalyst for improvements. While the evaluator observes, she/he ad vises on the emergency planning and operational process and fosters stakeholders' involvement.

As a result, during the RTE process stakeholders jointly define what, how, and who can improve the overall response, clearly outlining roles and responsibilities. As in developmental evaluation, the evaluator can become part of a design team helping to monitor what's happening—both processes and outcomes—in an evolving, rapidly changing environment of constant and immediate feedback (Patton, 1994, 2011). However RTEs still call for external report and accountability while development evaluations do not.

OPPORTUNITIES, CHALLENGES, AND LIMITATIONS

Even if they are designed to help managers be more effective and efficient while carrying out their difficult tasks, there are some constraints that have recurrently challenged the full uptake of RTEs. Taken positively, or "seeing the glass half full," these constraints represent opportunities.

1. To date the majority of RTEs have been rolled out in slow- and fast-onset "natural" disaster settings while very few have been fielded in complex and protracted emergencies.
2. Generally managers are primarily drained by the heavy workload on humanitarian staff dealing with the response, which leaves them with little time to engage properly with the evaluation team. As a result managers tend to have limited ownership of evaluation processes where evaluators are flown in and out.

3. Despite efforts made in recent years, baselines and other analytical documentation are missing at the early stage of a response.
4. Often fieldwork needs to begin quickly to be most useful while it takes time to identify available evaluators with the skills needed.
5. The necessary time for the evaluation team to do proper analysis is limited (Brusset et al., 2010).
6. As observed in Haiti and nowadays in the Philippines, the number of RTEs fielded in a specific crisis is skyrocketing, triggering so-called "evaluation fatigue."
7. National governments of the affected state are seldom involved in the evaluation process.
8. Funding necessary to field RTEs has not always been available.

Following major natural disasters such the 2004 Indian Ocean tsunami or the 2010 Haiti earthquake and in disaster hotspots such as Pakistan, RTEs have not been a one-off event. RTEs have been iterative, either fielded in different phases of the humanitarian response (emergency and recovery) to the crises or when a country experienced disasters of different types (e.g., Pakistan, 2005 earthquake and 2007 and 2010 floods). Yet, only two interagency RTEs have been fielded to evaluate the overall humanitarian response in complex emergencies, namely in the FATA (Federal Administered Tribal Areas and the Swat valley) and Swat valleys in Pakistan, and in 2005 in Darfur, Sudan. In open conflict situations, RTEs are rare due to serious security and access constraints. Despite having the necessary mechanisms in place, an RTE has not yet been fielded in Syria. So far only UNHCR and the Disaster Emergency Committee have fielded RTEs to look at the regional influx of refugees in host countries.

RTEs are one of the most challenging types of evaluations because teams are usually fielded within six weeks to six months following a disaster, when agencies are trying to scale up or already scaling down activities. The interagency RTE carried out in Haiti in 2010 (Grunewald, Binder, & Georges, 2010) was deployed just three months after the earthquake struck. However, in these circumstances the RTE can become burdensome to the agencies involved, and the exercise can suddenly become a "wrong time" evaluation. To tackle this risk, evaluation teams must be small and flexible with a very light footprint in the field because the entire team must fit in a Land Cruiser.

RTEs also have to be carried out within relatively short periods, and findings must be made available quickly. In general, teams have only two to three weeks to conduct the analysis and make the evaluation judgement before leaving the field. Nevertheless, as findings are then fed back for immediate use, RTEs can potentially identify and suggest solutions to operational problems as they occur and influence decisions when they are being made by feeding back aid recipients' and providers' views. Nowadays there is a growing tendency to describe any humanitarian evaluation as "real-time."

When fielded too late, after the disaster emergency response is over, the relevance of and need for an RTE could be questioned. During the first 10 months

following the earthquake in Haiti, 10 separate RTEs (for donors, the Red Cross, UN agencies, NGOs, and at interagency level) were fielded, which represented an enormous burden on staff and key informants and triggered evaluation fatigue on the ground. Agencies simultaneously initiating these RTEs claim that they have individual learning and accountability needs, but there is no evidence to suggest that the added value outweighs the costs.

Joint RTEs, in contrast, can add value through providing a mechanism for increased peer-to-peer accountability, particularly if the Humanitarian Country Team implements recommendations. By involving aid beneficiaries and local authorities, joint RTEs can also reinforce peer-to-peer and downward accountability and learning. However, such conditions can be difficult to achieve; so far, few have managed to involve the government involved at both central and provincial level.

Another challenge concerns who initiates and owns the evaluation. If headquarters initiate the evaluation, key stakeholders in the field are likely to be less involved in the identification of issues and key questions, or during implementation on the ground. For the evaluator, the challenge becomes understanding who poses the key questions and who will use the evaluation findings and recommendations. In only a few cases so far have management matrixes been drawn up (defining which recommendations had been accepted, who was responsible for taking action and implementing them, and what the deadline was for doing so) and acted upon after the reports were released.

THE INTERAGENCY RTE OF THE 2010 PAKISTAN FLOODS

At the core of an RTE is the idea that stakeholders—particularly those at country level—do not have to wait for the evaluation report to come out to start implementing the recommendations.

During the interagency RTE of the 2010 Pakistan floods, three provincial workshops and one national with key stakeholders were held as the main feedback mechanism. Findings, conclusions, and recommendations were initially presented by the team leader during the workshops. Stakeholders then jointly framed, validated, and prioritized the recommendations and defined the organization(s) responsible for implementing them (by whom) and timelines (by when). This process contributed to boosting ownership, uptake, and immediate implementation of the evaluation recommendations, fostering real-time learning and accountability among stakeholders engaged (Polastro, Nagrah, Steen, & Zafar, 2011).

The robustness of evidence presented in the draft report combined with bottom-up country workshops influenced immediate use at field level. HQ workshops, combined with regular presentations of lessons learnt on recent RTEs and meta-evaluations, contributed to the revision of policies and longer term use.

Another recurrent problem in many types of evaluations is the limited time available for consultations with beneficiaries. Careful planning ensures that time spent in the field ensures maximum stakeholder consultation. For instance, in Mozambique the RTE was both initiated and supported by the field, and four of the five team members could travel extensively to interview over than 400 beneficiaries at 16 different sites and observe the response in the four provinces affected by the cyclone and floods (Cosgrave et al., 2007). Similarly, in the 2010 Pakistan floods interagency RTE, the team incorporated lessons from previous evaluations and dedicated 80% of its time to field consultations visiting 20 site visits.

The lack of experienced evaluators is another key challenge; suitable candidates are generally booked three to six months in advance. Another limitation can be funding. In mid-2010 OCHA launched a call for proposals to field an interagency RTE in Kyrgyzstan, but no funding was secured; the Flash Appeal was also underfunded due to the time of year and the focus on other emergencies such as Haiti and Pakistan. In the case of the Pakistan displacement interagency RTE, it took time before donor funding was secured.

CONCLUSION

RTEs are not a new evaluation theory, but they represent an innovative way to evaluate humanitarian action. Today, when properly designed and led, they are a primary practice for improved learning, accountability, and performance across the sector. RTEs have grown as they have addressed specific stakeholders' needs and may still represent a better value for money as compared to more traditional ex-post evaluations. In humanitarian settings the demand for RTEs is rising, as the context on the ground tends to be fluid and goals are emergent and fast-changing rather than predetermined and fixed.

To consolidate their uptake, evaluation teams should identify who are the end users of the evaluation and specifically understand who needs what information and for what purpose and by when. Teams should systematically triangulate evidence drawing on a wide range of data sources. When findings and recommendations have a solid evidence base and have been properly validated, they will continue to bolster ownership and adaptive planning approaches that are responsive to contextual change.

When ownership exists, recommendations can immediately be used to address operational and organizational problems and in the longer term influence policy development. Overall, RTEs can make a significant contribution to real-time learning and improved efficiency and effectiveness within the humanitarian system.

That said, there is a risk that RTEs may become just a wasteful and ritualistic exercise, especially if the humanitarian sector does not understand when it is appropriate or not to launch them or how to launch them. As noted above, if RTEs are carried out too late or start overwhelming the field they may have an adverse effect, causing an "evaluation fatigue" that hampers utilization and uptake. The tendency to use them primarily for upward donor accountability purposes rather

than for field-level peer learning may rapidly undermine the added value of RTEs for personnel involved in the response. RTEs are most influential when carefully planned for and when done at the right time.

To maximize the potential contribution of RTEs to lesson learning, it is key that they adopt user-oriented approaches and become exercises that are "owned" and used by the field, rather than headquarters-imposed exercises. When possible, RTEs should be iterative and not carried out by flown-in evaluators who come and then disappear. Evaluators should facilitate reflection in different critical phases of the response.

Overall, to improve the humanitarian system's planning and performance, RTEs should be done jointly and at the right time. A triggering mechanism, as outlined in the transformative agenda and in several agencies' evaluation policies, to ensure that RTEs happen and that adequate human and financial resources are allocated must be more systematically used even in those situations such as Syria where access and security may be limited.

Incentives for improving knowledge management and fostering real-time learning and accountability should be identified at the field level. The first step would be to regularly roll out bottom-up workshops with key stakeholders to validate and prioritize recommendations presented in the draft report and assign responsibility for implementation. A second step would be to systematically define who is responsible for implementing recommendations produced and draw action plans once these have been formulated.[12]

Once the action plans are agreed upon, they should be regularly monitored. In addition, to ensure adequate involvement of field-level stakeholders in the RTE (including aid recipients and local authorities, when possible), initiating organizations need to provide regular feedback to them on the implementation of recommendations. Last but not least, lessons from RTEs should be better disseminated to the broader evaluation community.

ACKNOWLEDGEMENTS

NOTES

1 According to its Latin etymology, "to evaluate" means "to ascertain the value or worth of."
2 ALNAP membership works together to identify common approaches to improved performance; explore new ways to improve learning and accountability through

evaluations and other mechanisms; and share ideas, lessons, and innovations. The overall aim is to make an active contribution to solving longstanding challenges facing the sector. For more information see the minutes of the ALNAP meeting in Nairobi and the record of the RTE workshop under the auspices of UNICEF and OCHA at the ALNAP meeting in Rome, www.alnap.org/meetings/pdfs/20meetingreport.pdf

3 According to Patton (1994, p. 317), "Utilization-focused evaluation is a process for making decisions about and focusing an evaluation on intended use by intended users. Utilization-focused evaluation shifts attention from methods or the object of evaluation (e.g., a program) to the intended users of evaluative processes and information, and their intended uses. The evaluator, rather than acting as an independent judge, becomes a facilitator of evaluative decision-making by intended users. Thus, the utilization-focused evaluator facilitates judgements about merit and worth by intended users rather than highlighting his or her own judgements."

4 E.g., Real-Time Evaluation of Norway's International Climate and Forest Initiative, http://www.oecd.org/derec/norway/48086385.pdf

5 Performance indicators are measures of inputs, processes, outputs, outcomes, and impact for development projects, programs, or strategies. When supported with sound data collection (perhaps involving formal surveys), analysis, and reporting, indicators enable managers to track progress, demonstrate results, and take corrective action to improve service delivery (World Bank, 2004).

6 Using Michael Scrivens' distinction (1991) when he quoted Robert Stake: "When the cook (or chef) tastes the soup, that's formative; when the guests taste the soup, that's summative." Building on this distinction, we can liken RTE to a team of sous chefs that often involve the restaurant guests in the kitchen (e.g., senior management, decision-makers, communication experts, external audiences, members of the affected populations) to taste the soup before serving in the dining area. "If not to their palate, ingredients can be changed or possibly the entire effort discarded."

7 For more information on the transformative agenda, see http://www.humanitarian info.org/iasc/pageloader.aspx?page=content-template-default&bd=87

8 https://www.humanitarianresponse.info/news/publication-ia-rte-pakistan-floods-report

9 See more at http://daraint.org/2013/04/24/4481/alnap-a-call-for-evidence-based-decision-making-in-humanitarian-response/#sthash.rczXKmLo.dpuf

10 A snowball sample is a sample that is expanded by asking the initial interviewees to recommend others they believe would be useful sources of information. The process begins by asking them, "Who knows a lot about …? Whom should I talk to ….?" In snowball sampling, you begin by identifying someone who meets the criteria for inclusion in your study. You then ask them to recommend others who they may know who also meet the criteria. Although this method would hardly lead to representative samples, there are times when it may be the best method available. Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find. See http://www.socialresearchmethods.net/kb/sampnon.php

11 See the link to the short video used to disseminate both the findings and the ownership of the process during the Inter Agency Real Time Evaluation of the Humanitarian Response to 2010 floods in Pakistan. http://www.youtube.com/watch?v=qW-XR14k4NQ

12 About two thirds of the OECD DAC bilateral donors have mechanisms in place to ensure that management responds to and follows up on evaluation findings (OECD DAC, 2013).

REFERENCES

Bamberger, M. (2008). Enhancing the utilization of evaluations for evidence-based policy making. In M. Segone (Ed.), *Bridging the gap: The role of monitoring and evaluation in evidence-based policy making* (pp. 120–142). Geneva, Switzerland: UNICEF.

Beck, T., & Buchanan-Smith, M. (2008). *Joint evaluations coming of age? The quality and future scope of joint evaluations.* London, UK: ALNAP. Retrieved from http://www.alnap.org/pool/files/7rha-Ch3.pdf

Brusset, E., Cosgrave, J., & MacDonald, W. (2010). Real-time evaluation in humanitarian emergencies. In L. A. Ritchie & W. MacDonald (Eds.), *Enhancing disaster and emergency preparedness, response, and recovery through evaluation, New Directions for Evaluation,* 126, 9–20.

Buchanan-Smith, M., & Cosgrave, J. (2013). *Evaluation of humanitarian action: Pilot guide.* London, UK: ALNAP. Retrieved from http://www.alnap.org/resource/8229

Carlsson, J. (2000) Learning from evaluations. In J. Carlsson and L. Wohlgemuth (Eds.), *Learning in development co-operation* (pp. 120–129). Stockholm, Sweden: Almqvist & Wiksell International.

Cosgrave, J., Gonçalves, C., Martyrs, D., Polastro R., & Sikumba-Dils, M. (2007). *Inter-agency real-time evaluation of the response to the February 2007 floods and cyclone in Mozambique.* New York, NY: IASC. Retrieved from https://ochanet.unocha.org/p/Documents/MZ_RTE_Report_Final%5B1%5D.pdf

Cosgrave, J., Polastro, R., & Zafar, F. (2010). *Inter-agency real time evaluation (IA RTE) of the humanitarian response to Pakistan's 2009 displacement crisis.* New York, NY: IASC. Retrieved from http://daraint.org/wp-content/uploads/2010/10/IARTE_PK_displacement_2010_report.pdf

Cosgrave, J., Ramalingam, B., & Beck, T. (2009). *An ALNAP guide pilot version.* London, UK: ALNAP. Retrieved from http://www.alnap.org/pool/files/rteguide.pdf

Feinstein, O., & Beck, T. (2006). Evaluation of development interventions and humanitarian action. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 536–558). London, UK: Sage.

Good Humanitarian Donorship Initiative. (2003). *Principles and good practice of humanitarian donorship.* Stockholm, Sweden: Author. Retrieved from http://www.goodhumanitariandonorship.org/gns/principles-good-practice-ghd/overview.aspx

Grunewald, F., Binder, A., & Georges, Y. (2010). *Inter-agency real-time evaluation in Haiti: 3 months after the earthquake.* New York, NY: IASC. Retrieved from www.alnap.org/pool/files/haiti-ia-rte-1-final-report-en.pdf

Hallam, A. (1998). *Evaluating humanitarian assistance programmes in complex emergencies: Good practice review.* London, UK: ODI HPN. Retrieved from http://www.odihpn.org/index.php?option=com_k2&view=item&layout=item&id=2109

Herson, M., & Mitchell, J. (2005). Real-time evaluation: Where does its value lie? *Humanitarian Exchange Magazine, 32.* Retrieved from http://www.odihpn.org/humanitarian-exchange-magazine/issue-32/real-time-evaluation-where-does-its-value-lie

Jamal, A., & Crisp, J. (2002). *Real-time humanitarian evaluations: Some frequently asked questions* (EPAU /2002/05). Geneva, Switzerland: UNHCR Evaluation and Policy Unit. Retrieved from http://www.unhcr.org/research/RESEARCH/3ce372204.pdf

Nutley, S., Walter, I., & Davies, H. (2007). *Using evidence: How can research inform public services*. Bristol, UK: Policy Press.

OECD DAC. (1999). *Guidance for evaluating humanitarian assistance in complex emergencies* (DAC Evaluation Series). Paris, France: OECD. Retrieved from http://www.oecd.org/development/evaluation/2667294.pdf

OECD DAC. (2006). *Guidance for managing joint evaluations* (DAC Evaluation Series). Paris, France: OECD. Retrieved from http://www.oecd.org/development/evaluation/37512030.pdf

OECD DAC. (2013). *Evaluating development activities: 12 lessons from the OECD*. Paris, France: OECD. Retrieved from http://www.oecd.org/dac/peer-reviews/12%20Less%20eval%20web%20pdf.pdf

Patton, M. Q. (1994). Development evaluation. *Evaluation Practice*, *15*(3), 311–319. Retrieved from http://dx.doi.org/10.1016/0886-1633(94)90026-4

Patton, M. Q. (2008). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.

Polastro, R. (2011). Real time evaluations: Contributing to humanitarian system-wide learning and accountability. *Humanitarian Exchange Magazine, 52*.

Polastro, R., Khalif, M. A., Van Eyben, M. N., Posada, S., Mohamoud Salah, A. S., Steen, N., & Toft, E. (2011). *IASC evaluation of the humanitarian response in south central Somalia 2005–2010*. Nairobi, Kenya: HCT. Retrieved from www.oecd.org/redirect/dataoecd/9/29/49335639.pdf

Polastro, R., Nagrah, A., Steen, N., & Zafar, F. (2011). *Inter-agency real time evaluation of the humanitarian response to Pakistan's 2010 flood crisis*. New York, NY: IASC. Retrieved from http.www.alnap.org/pool/files/1266.pdf

Poole, L., Walmsley, L., Malerba, D., Sparks, D., Sweeney, H., Smith, K., . . ., & Coppard, D. (2012). *Global humanitarian assistance report 2012*. Wells, UK: Development Initiatives, London.

Sandison, P. (2003). Desk review of real-time evaluation experience (UNICEF Evaluation working paper). New York, NY: UNICEF.

Sandison, P. (2006). The utilisation of evaluations. In J. Mitchel (Ed.), *ALNAP review of humanitarian action: Evaluation utilization* (pp. 89–144). London, UK: ALNAP. Retrieved from www.alnap.org/pool/files/ch3-f1.pdf

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Telford, J. (2009). *Joint review of experience with IASC-mandated real-time evaluations and the future of joint humanitarian evaluations*. New York, NY: IASC.

World Bank. (2004). Monitoring and evaluation: Some tools, methods and approaches. Washington, DC: Author.

AUTHOR INFORMATION

Riccardo Polastro is a p rincipal consultant at IOD PARC, an international UK-based development consultancy. He has mo re than two decades of management experience in humanitarian affairs and development aid, having worked in over 65 countries for the International Movement of the Red Cross, the United Nations, NGOs, and donor agencies. In his p revious role as t he head of evaluation at DARA (2006–2013), Ricca rdo worked extensively on policy and operational evaluations. He is a former member of the ALNAP steering committee. He also regularly lectured in several universities' MA p rograms and provided professional training.