Journal of MultiDisciplinary Evaluation
Volume 10, Issue 23, 2014

ISSN 1556-8180
http://www.jmde.com

Real-Time Evaluation of Humanitarian Assistance Revisited: Lessons Learned and the Way Forward

Susanna Krueger
Global Public Policy Institute (GPPi)

Elias Sagmeister
Global Public Policy Institute (GPPi)

Background: The real-time evaluation (RTE) approach has been applied in humanitarian assistance for two decades. Its spread seems to be as much a result of entrepreneurial evaluators who aim to demonstrate expertise to potential clients as it is a response to actual demand for immediate feedback of results to decision makers.

Purpose: As RTE has come under scrutiny recently, this study demystifies the concept of RTE and looks beyond textbook descriptions of its advantages in order to understand its practical application in humanitarian action.

Setting: NA

Intervention: NA

Research Design: NA

Data Collection and Analysis: NA

Findings: It then suggests lessons for how to improve the application of real-time evaluations and related concepts in practice.

Keywords: *real-time evaluation; humanitarian assistance; organizational learning; complexity science.*

Introduction

The concept of real-time evaluation (RTE) has evolved and spread considerably since its introduction some two decades ago, and since reviews first looked comprehensively at its application (Herson & Mitchell, 2005; Sandison, 2006). It has formed an integral part of evaluation practice among humanitarian organizations, who are continuing to further develop similar approaches for immediate evaluative feedback. The most recent example is is the case of the emerging Inter-Agency Standing Committee concept of Real-Time *Operational Reviews* (see also p. 5). These organizations face manifold challenges with regard to access to populations, lack of quality data, rapidly changing agendas and demands from policymakers in a generally chaotic (that is, unpredictable) crisis situation. While struggling with these challenges, they need to show results and the ability to learn if they want to sustain support from donors and the wider public – and make a difference on the ground. Real-time evaluation should support them in this difficult task.

However, the spread of "real-time" approaches seems to be as much a result of entrepreneurial evaluators who aim to demonstrate expertise and innovation to potential clients as it is a response to actual demand. Today, RTE is often presented as the newest and sharpest tool in the evaluator's toolbox. This article will first turn to broader trends in evaluation theory and debate, such as evaluation under complexity and its potential to stimulate organizational learning, to help understand where real-time evaluation came from and why it is so popular. Our study then uses a systematic review of evaluation reports (see Appendices A and B for a list of reports covered and interviews conducted) and key informant interviews to see the extent to which current practice is congruent with the popular theoretical descriptions of RTE. By looking at the reports of practical application of RTE in humanitarian action, the study demystifies the approach and takes stock of its strengths and weaknesses. Finally, we suggest in what situations real-time evaluation makes most sense and how it could be improved.

We argue that in order to improve the practical benefit of real-time evaluation, humanitarian organizations need to be even more selective and modest in its use. Wherever real-time evaluation is used – it should prioritize endogenous learning in organizations over questions of accountability and control.

Method

The study is based on a meta-evaluation of 44 RTE reports. A caveat is in order here: We used purposive sampling to select publicly accessible evaluation reports from influential aid agencies. While our sample is thus not representative of the whole population of real-time evaluations ever conducted, it shows major trends among key aid actors. Compared to other domains of evaluation, methodological standards appear more modest in the humanitarian context. This meta-evaluation did therefore not entail all the standard quality criteria, but is a somewhat more basic assessment of evaluation practice compared to the promises of the Real-Time evaluation approach. While the analysis thus falls short of extensive meta-evaluations as demonstrated by other authors (Scriven, 1969, 2008; Daniel Stufflebeam, 2001), it aims to provide an objective but realistic assessment through a number of criteria derived from the description of RTE (see chapter 3), as illustrated in Table 1. To assess the congruence of RTE practice with formal definitions and approaches described by major stakeholders, we thus looked at formal aspects, management aspects, the methods applied, the content and scope of RTE and whether reports identified general lessons or showed "good practice" worth highlighting.

RTE is arguably more about the monitoring and evaluation process than about the final report. We thus supplemented our desk review with seven expert interviews that have experienced this process and could speak to how RTE is used in specific organizations, the role of RTE in their broader evaluation policy and the experience with the approach so far. We used purposive sampling to include aid practitioners and third parties (consultants) with experience in real-time evaluations. Our assessment is additionally informed by practical evaluation experience of the authors.

Table 1
Key Criteria for Analysis

Assessment Categories	
Formal aspects	• Organization • Sector and country • Point of time in project cycle • Duration of RTE
Management	• Trigger for RTE • Team composition • Communication of results • Follow-up mechanism
Methodology	• Sampling and data collection • Use of technology • Participation of beneficiaries • Design option
Content and scope	• Objectives and rationale • Use of evaluation criteria (e.g. DAC, n.d.)

The RTE Approach, Relevant Trends and Theory

In some shape or form, the concept of RTE has likely been around for more than two decades (see, for example, Hobday (1988)). Its breakthrough in the humanitarian sector is largely attributed to the United Nations Refugee Agency's (UNHCR) initial proposal of something akin to RTE in 1992 during an evaluation of its response to the Persian Gulf crisis (Crisp, Martin, & Prattley, 1992). In subsequent documents, UNHCR summarized the perceived advantages of RTE in three words: (a) timeliness, (b) perspective and (c) interactivity (Jamal & Crisp, 2002). Having conducted RTE in more than 10 countries in the past decade, UNHCR's Evaluation and Policy Analysis Unit (EPAU) has pushed and shaped the methodology considerably. For example, UNHCR has conducted RTE in the Sudan-Eritrea emergency (2000), Angola (2000), or Afghanistan/Iran (2002). More recently, a multi-country review of the cluster approach in Chad, DRC, Uganda, Somalia, Liberia (2007), and Sri Lanka (2010) was completed under UNHCR's lead.

Today, the term is used in the evaluation policies and guidelines of many humanitarian and development organizations. As Table 2 illustrates, the definitions held by other major humanitarian and development organizations largely agree on the same advantages. Oxfam, CARE and other NGOs have all adapted very similar definitions (OXFAM, n.d.). An apparent difference lies in the inclusion of impacts in the original UNHCR definition.

RTE approaches thus have to cater to great expectations and tend to be described as an all-in-one evaluation solution. A case in point is the description UN OCHA gives in their terms of reference: "The applied methods for IA RTE shall be light and participatory but rigorous as well" (UN OCHA, 2011, p. 3).

Existing reviews (Herson & Mitchell, 2005; Sandison, 2003) agree on the following features of the RTE approach:

- RTE usually takes place while implementation of the respective program is still going on, thus allowing immediate changes;
- RTE is implemented in a fast and flexible way, typically taking only a few days instead of weeks;
- The focus of RTE is on processes instead of outcomes or even impacts, sometimes with limited regard for standard evaluation criteria (e.g. DAC, ALNAP);
- Ideally, more than one RTE is conducted in an iterative, on-going evaluation process to optimize respective programs;
- Compared to classic ex-post or formative evaluations, internal staff has a much more important role to play also in the research design, with external consultants facilitating the process.

International humanitarian bodies and networks, such as the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) and the UN Inter Agency Standing Committee (IASC), have refined the concept in their own guidance and adapted it to their work. According to the 2009 ALNAP guide for RTE (Cosgrave, Ramalingam, & Beck, 2009), what distinguishes real-time from regular evaluations is the timeframe and the audience of these studies. Contrary to "classical" evaluation studies, an RTE should deliver its findings before evaluation staff leave the field with the primary audience being the implementing agency's staff. As far as the timeframe is concerned, RTE should "look at today to influence tomorrows planning" (p. 11). Methods suggested include key informant and beneficiary interviews, data analysis, observation and document review (p. 57).

Finally, joint RTE involving a number of humanitarian actors have been piloted and UN agencies active in humanitarian assistance developed a set of operating procedures that specify criteria for triggering RTE, intended purpose and standard terms of reference (IA RTE Support Group, 2010). Inter-agency RTE are characterized by their shared management and methodological oversight through interagency reference and management groups. Again, expectations are high as guidance describes the approach as light, agile approaches, with restricted scope, and participatory methods, with staff of agencies responding to crisis in the respective countries as their main intended users (IA RTE Support Group, 2010, p. 5). More recently, under its new "Transformative Agenda", the Inter Agency Standing Committee (IASC), an inter-agency forum of UN and non-UN humanitarian organizations, has announced the development of Real-Time Operational Reviews. According to a IASC reference document (IASC, 2012, p. 5), "this will be based in part on the concept of the Real-Time Evaluation (RTE), but place much greater emphasis on the validity of the strategic response plan and how the system is performing against agreed targets. The RTOR will focus on "why" any response targets are not being met, with corresponding immediate and medium term recommendations to overcome challenges." It remains to be seen how these operational reviews will differ from existing RTEs in practice.

Table 2
Definitions of Real-time Evaluation

UNICEF	UNHCR	DG ECHO
"A Real-time Evaluation is one carried out whilst a programme is in full implementation and almost simultaneously feeds back its findings to the programme for immediate use." (Sandison, 2003, p. 8)	"A Real-time Evaluation (RTE) is a timely, rapid and interactive peer review of a fast evolving humanitarian operation, undertaken at an early phase. (…) The broad purpose is to gauge the effectiveness and impact of a given UNHCR response, (usually in an emergency) and to ensure that its findings are used as an immediate catalyst for operational change." (Jamal & Crisp, 2002, p. 1)	"RTE is a very fast and timely means of providing evaluative feedback. It is launched sufficiently early in the course of a humanitarian intervention to have an effect on key decisions and to assist in improving the existing operation." (DG ECHO, 2007, p. 36)

Embedding RTE in Recent Organizational Development and Evaluation Trends

The theoretical characteristics of RTE outlined above show that the main focus is on improving projects and programs as they are implemented. These features mirror three broader trends in evaluation and organizational development that have shaped evaluation practice and the expectation of humanitarian policymakers in past decades: an overall culture of *urgency*, the growing need for *adaptive learning* and increasing *recognition of complexity*. As we will see, these trends paved the way for RTE's triumphal procession into evaluation practice today and go a long way in explaining its popularity.

Evaluation Under Urgency and Resource Constraints

In humanitarian assistance, evaluation had a fairly late start compared to development assistance and only really took-off after the traumatizing experience of collective failure in Rwanda in the mid-nineties. Donors (and implementers) have since then realized that the sector as such is more potent than ever before and that the risk of unintended harm requires greater scrutiny (Collinson & Elhawary, 2012). More funds and more competition among a growing number of humanitarian actors have also increased the pressure to demonstrate results, accountability and learning. A "culture of evaluation" in humanitarian assistance thus emerged, largely influenced by development evaluation (Crisp, 2000). However, the sector showed a preference for less rigorous and shorter term methods, assuming that in the face of dramatic suffering and rapid need for action, one cannot afford to evaluate results of programs in great depth.

The "culture of evaluation" thus met with a "culture of urgency" and quick, cheap, and useful methods had to be found to satisfy the emerging demand for accountability and to legitimate the resources allocated to humanitarian action. Increasing professionalization and more funds devoted to innovative ways of M&E set the stage for RTE, which promised to satisfy the demand while respecting the limits of time and resources available.

Evaluation to Facilitate Adaptive Learning

In the humanitarian sector, the pendulum has swung from accountability (especially "upwards" accountability towards donors) towards learning purposes of the evaluation activity. Hence, following the multi-agency evaluation of the international response to the Rwanda genocide in 1997, the "Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP)" was established with the purpose to improve humanitarian performance through *improving the quality and utilization of evaluations* and supporting *organizational and system-wide change*. (In 2000, after the Kosovo crisis, ALNAP also developed the concept of field based "Learning Offices" that should allow organizations to share lessons and learn "in real time" as a crisis evolves. After a series of pilot studies, the concept was not further developed.)

The growing demand for learning was fuelled and accompanied by frequent observations in the development and humanitarian sector that results from evaluations were not used for organizational change and that too many lessons came too late (Patton, 1997; Sandison, 2006; van de Putte, 2001; Weiss, 1977). Clark & Ramalingam (2008) point out that "[…] work on field-level learning has also shown that, despite good intentions and sound starting points, many initiatives focus on documents, systems and products, and so fail to support learning in operational contexts, which is generally a social, human and tacit process."

Evaluation was thus increasingly expected to provide a continuous stream of information that helps managers to reach effective decisions and improve programs in the long run (Patton, 2011; Stufflebeam, 1983). In other words, evaluation's most important purpose was not seen to be *proving*, but *improving (Stufflebeam, 2004)*. The evolving understanding is perhaps best illustrated by Patton's "developmental evaluation", which should help "[d]evelop a rapid response in the face of a sudden major change or a crisis, like a natural disaster or financial meltdown, exploring real-time solutions and generating helpful interventions for those in need" (p. 21).

Despite all the emphasis on learning and use, a rather simplistic understanding of the learning subject (organizations) remained prevalent and the most frequent metaphor for organizations was still the machine (Clark & Ramalingam, 2008). So while use had become more important, the users (here: the actors in the humanitarian contexts) were still not fully understood. Like the logical program theories, evaluators expected evaluation to "cause" learning, when in fact evaluation influence is less direct and automatic and often depends on many different processes (Mark & Henry, 2004, p. 48).

In the field of organizational development though, influential authors (C. Argyris, 1990; Chris Argyris & Schön, 1978; Schein, 1969; Senge, 2006; Weick, 1995) had long pointed to the limits of predicting organizational learning. They therefore developed approaches to facilitate organizational learning, which are informed by a constructivist world view (Lewin, 1946).

Once these thoughts gained acceptance from managers and policymakers and organizations were increasingly seen as complex systems, change could not simply be "induced", but needed to be "facilitated" through the participation of responsible stakeholders. Instead of expecting outsiders (i.e. external evaluators) to trigger a learning process through expert analysis and a long list of recommendations, organizations recognized the need to build internal evaluation and learning capacities. RTE, as described above, in theory incorporates these considerations and offers to enhance learning by organizations.

Complexity and evaluation. Weaver (1948) already distinguished between *simple* phenomena which show predictable interactions of limited parts, *complicated* phenomena where multiple but stable parts interact, and *organized complexity*, where multiple parts interact in unpredictable, yet patterned ways. (Other similar categorizations and heuristics have been developed by (Ackoff, 1999) and others. (Mitchell, 2009) assumes about 40 different measures of complexity are circulating in the literature.) Similarly, Snowden and colleagues stress that a situation is neither just simple nor complex, but that different aspects are found in every situation, requiring different decision strategies. Kurtz and Snowden (2003) thus distinguished between simple, complicated, complex and chaotic aspects of situations and corresponding decision making models.

Needless to say, humanitarian assistance usually involves multiple actors who come together in dynamic crisis situations characterized by non-simple phenomena. The traditional view of evaluation as a rather static endeavour in which an agreed-upon theory of change and predefined indicators of success are analysed in a formative or summative manner was thus challenged in the humanitarian field as it was elsewhere (Patton, 1997; White, 2009). Increasingly, practitioners and scholars agree that this understanding of evaluation and corresponding practice is at odds with the complexities of working with social systems, and particularly humanitarian action. This trend is perhaps best illustrated by recent conferences on topics such as "Systemic Approaches in Evaluation" (GIZ 2011), "Navigating Complexity" (Wageningen 2008/09), "Evaluation in a Complex World" (EERS conference, 2012), or "Evaluation in Complex Ecologies" (AEA 2012).

Hence, innovative evaluation approaches have to show they can address this complexity and are up for informing decision making and learning. Most evaluators agree that for learning and feedback loops in complex and uncertain environments to occur, the *process* of evaluation is more important than formal compliance with rigid logical frameworks (Ling, 2012; Pawson & Tilley, 1997).

Some 50 years after Weaver's distinction, the rigid planning and requirement to demonstrate results in bureaucratic aid organizations shows that there is still a long way to go. Nevertheless, complexity-sensitive approaches to aid management and evaluation have experienced a rise and are more popular today than ever (Jones, 2011; Ramalingam, Jones, Reba, & Young, 2008).

The RTE approach not only corresponds with these trends, but offers to bridge the seemingly contradictory demands for complexity-sensitive approaches that can foster adaptive learning in organizations, with a light footprint on the resources of donors and implementers – in real-time. The remaining part of the paper will look at how the approach has lived up to these expectations in practice.

How is RTE Applied in Practice?

To compare the theory of the RTE approach with its actual application in humanitarian organizations, we analyzed a sample of 44 publicly accessible RTE implemented between 2000 and 2010 and documentation from nine large aid organizations. All but four reports were hence conducted after the last major desk review was undertaken by Sandison (2003). Table 3 provides an overview of the reports considered and countries covered in these evaluations. As noted above, we selected publicly accessible evaluation reports purposively to cover the work of influential aid agencies that arguably influence evaluation practice of the broader sector. We analyzed reports using a template that covered the assessment criteria described in Table 1 above. The full table used for this analysis is available from the authors.

Table 3
Overview of Reports Considered by Organization, Countries, and Regions

Organization	Countries and Regions Covered	Number
CARE	Sudan	2
CRS	Pakistan, Haiti	3
DEC	Pakistan	1
DG ECHO	Zimbabwe, Haiti	2
FAO	Global, Indonesia, Sri Lanka, Thailand, Maldives	3
HAP	Afghanistan, Sierra Leone	2
IASC	Pakistan, The Philippines, Sudan, Horn of Africa, Mozambique, Haiti	10
IFRC	Indonesia and Sri Lanka, Asia (India, Sri Lanka, Maldives) and East Africa (Somalia, Seychelles)	2
Oxfam	Zimbabwe, The Philippines, Laos, Cambodia, Indonesia	4
Tearfund	Haiti	1
UNHCR	Afghanistan, Angola, Chad, DRC, Liberia, Somalia, Sri Lanka, Sudan, Eritrea, Uganda	10
UNICEF	Yemen, Georgia	2
WFP	Zambia, Zimbabwe, Swaziland, Mozambique, Malawi, Lesotho, India	2
		44

The analysis shows that while theoretical definitions held by organizations are largely congruent, implementation of the concept varies considerably. Differences in timing, team composition and scope of the evaluation are most apparent. Clearly, the title “RTE” is used more broadly than its original conceptualization would justify.

Regarding their position in the project life cycle, we see that RTE is not only used at the beginning of projects. While most were conducted in the early phases, i.e. 2-4 months into the response (13); many were conducted 5-6 months into response (6) or much later, 1-3 years in (7). Despite the aim to influence on-going programming and implement immediate changes, about one third of the RTE considered for this review have been implemented towards the end of programs.

Frequently, timing was affected by delays in the recruitment of evaluation teams or the conceptualization of studies, despite plans to initiate the evaluation earlier. What was planned as a real-time exercise to influence on-going operations was sometimes implemented much later, leading to critique of “wrong time evaluations” (IASC Mozambique, 2007). In other cases, evaluation teams adjusted their focus accordingly, to then generate lessons for future endeavours.

The analysis showed a large variation in duration, from one week to 1.5 years, with the majority taking between 4 and 15 weeks. Thus, most of the studies took longer than would be expected from the literature. (For example, while UNHCR completed very quick studies, FAO’s RTE went through a very long process of 1.5 years.) In many cases, the exact duration was not apparent from the reports, but it seems that only in exceptional cases did RTE “look at today to influence tomorrows planning,” as suggested in the ALNAP guide.

Most organizations emphasized the demand for learning as the main motivation for an RTE. However, most combined the learning objective with expectations of accountability. In about 10% of the cases, the RTE was motivated by some inadequacy, negative feedback or failure. Other reasons given were the transformation to another phase, exceptionally large or complex interventions, or drastic changes in the project or emergency context.

In terms of stated objectives, almost all RTE reports and available terms of references highlight the need for immediate feedback to improve on-going operations and to feed into future programming and strategy. Objectives given are often broken down into concrete questions and cover a diverse and broad range of strategic and operational issues. In contrast to the narrow focus prescribed to RTE, evaluation questions formulated in terms of references addressed longer-term strategic issues of broad scope.

Organizations show different strategies and expectations concerning team composition and responsibility for the evaluation process. CARE, FAO, Oxfam, UNHCR and WFP deployed teams of experts from their own organizations, sometimes drawing on selected additional support from external consultants. On the other hand, DG ECHO, IASC, IFRC, UNICEF, CRS, DEC relied on consultants to lead evaluation missions. Overall, the role of external consultants has in many cases been interpreted to go beyond that of a facilitator. Many organizations followed the familiar pattern of regular evaluations and expected mostly external advice from experts and evaluators.

The degree to which recommendations from RTE reports have been implemented cannot be judged from our desk review. However, some noteworthy examples included a follow-up workshop some months after completion of the evaluation (UNICEF, 2010) and publicised reactions to recommendations and management-response matrices (FAO, 2010b). Similarly, IASC (2006) and IFRC (2005a, b) included matrices in their reports or attached in separate documents. Usually, organizations held debriefings or workshops with key stakeholders from country offices and headquarters to feedback results and discuss emerging findings.

With the notable exception of the IASC RTE of the 2007 response in Mozambique, reports did not go into much detail when describing the methods applied. (As one interviewee pointed out, commissioning organizations sometimes even explicitly ask to cut “boring methodology sections” to produce short reports.) Sampling was generally done on an ad-hoc basis. Evaluators then mostly relied on paper and pencil, conducting individual, group and (telephone) interviews. Beyond telephone and email, teams did not make use of technology to collect data.

As for the design option applied, evaluation reports remained very vague in their description of the exact design chosen. Virtually none showed efforts to randomize sampling or apply (quasi-) experimental research setups.

More than two-thirds of RTE considered made some effort to include the perspective of beneficiaries. Anything from a dozen (UNICEF Georgia, 2009) to several hundred (IASC Mozambique, 2007) beneficiaries were included in analyses. Beneficiary feedback was largely seen to complete other perspectives, but was never an explicit focus. Documentation of where, when and how beneficiaries were included (e.g. through group or individual interviews, surveys or workshops) or how samples have been selected was either missing or incomplete in all but two reports (IASC Mozambique, 2007; IASC Haiti, 2010).

About one third of the reports considered mention DAC criteria for evaluation, but only selected criteria are then applied. Less frequently, Sphere standards, minimum standards in core areas of humanitarian assistance, are addressed (ECHO, 2007; IASC, 2007; IFRC 2005a) or reference to UNEG standards of evaluation (IASC, 2010; UNICEF, 2009) or ALNAP guidelines (DG ECHO, 2007; IASC, 2010) is made.

Summary of Results

The analysis shows that the practice of RTE diverges from the theoretical characterizations frequently described. The main differences are:

- Organizations use RTE for more than the specific niche they were intended for;
- They expect RTE to consider results and impacts, going beyond the narrower scope on processes;
- Humanitarian actors apply RTE along all phases of the project life cycle, not just at the beginning of projects;
- The time required to conduct an RTE is longer than expected in the literature;
- Evaluators are still expected to provide external expert advice on operational issues;
- Humanitarian organizations use RTE for accountability purposes as well as to facilitate learning.

This shows a current practice where RTE are not used to their full potential, and in some cases even misused for less adequate purposes.

Most significantly, an imbalance is evident between the light and agile setup theoretically described and the comprehensive questions major organizations pose in their terms of reference. The scope is frequently extremely broad, including wish lists of interesting aspects of an intervention from sector-relevant results to longer-term impact. This is problematic for at least two reasons:

First, RTE is not able to provide solid analysis about results and impact with the simplified and quick set-up they usually entail. The knowledge they generate is neither as reliable nor as valid as the findings generated by regular evaluations or impact studies with more elaborate designs. Moreover, findings from RTE cannot be generalized beyond a fairly specific context.

Second, being the more affordable option, they can crowd out full-fledged evaluation efforts, weakening the overall evaluation system in humanitarian assistance instead of merely supplementing it with a more agile approach.

But even if done correctly in accordance with the theoretical descriptions of the approach, RTE has its limits. Accepting these limits would lead to a more modest – and less formal approach. RTE should thus be seen as a helpful approach for organizational innovation and learning in complex environments rather than a (humanitarian aid) evaluation methodology. Thus, the application of RTE needs a number of corrections and the development of future real-time evaluative approaches could benefit from the following suggestions.

How Could RTE be More Useful?

1. Place Them in the Right Context and Accept The Limits

In light of extensive praise from existing (grey) literature and guidelines, the limits of RTE should not be overlooked. RTE does not provide an all-in-one solution, but is intended for a specific niche and potentially harmful if used for other purposes. Its agile and economic character comes at a price in rigor and depth. While rigor does not have to entail complicated methods or necessarily depends on set quality standards, it does take time and will slow down the evaluation process (Zelik, Patterson, & Woods, 2010).

The goal should be to help decision makers understand the situation they attempt to influence. RTE can help to learn more about how humanitarian action influences a specific context and vice versa, when original plans clash with reality. To find general lessons for all organizations and contexts, other approaches remain more expensive, but also more feasible.

Moreover, the context needs to be right. RTE is most appropriate in “complex” situations (cf. Kurtz & Snowden, 2003). In such contexts, staff in organizations cannot rely on their pre-determined plans to guide decision making, but will have to take action based on the limited information they have, sense available options and try new approaches. RTE can add value in providing staff of humanitarian organizations with the proper tools for "sensing" how the initial probing or acting works out in practice and what it means for the evolving response. RTE, done right, can thus "support informed adaptation by practitioners" before it is too late for the necessary changes (Ling, 2012, p. 84).

2. Do Not Use RTE to Scrutinize Implementing Organizations

In line with the limitations highlighted above, RTE should not be applied where concerns about accountability to donors outweigh or compromise learning purposes. For donors to judge whether their funds have been put to good use, longer-term standard evaluation with an external control function should remain the mechanism of choice. RTE, on the other hand, adds most value in a trusting and flexible environment, where mistakes can be used as opportunities to optimize processes before it is too late.

Donors need to accept the uncertainty and risks involved in evolving humanitarian action and encourage their partners to exploit all means to make sense of the complex situations they are in. They should clearly differentiate between internal learning mechanisms such as RTE and accountability or control-oriented evaluations that focus on results. The most important type of accountability that should be fostered along the way and to which the findings of RTE can contribute, is the *accountability for learning*. In other words, donors should encourage implementing organizations to set up effective internal learning mechanisms and to allow affected populations to influence programming as it develops (cf. point 4). They should not, however, compromise the learning environment by micro-evaluating processes in real time. As a consequence – and contrary to standard evaluations – the results from RTE should not automatically be made public, but rather used to their fullest inside the organization.

3. Use Them as an In-House Tool for Organizational Development and Learning With Staff on The Ground

As Eggers (2006) points out, planning and evaluation are two sides of the same coin. Today's RTE needs to be able to influence tomorrow's planning. To achieve this, the organizational setup needs to be adjusted accordingly. For example, organizations could involve staff that is responsible for planning in RTE and should not only delegate responsibility to outside experts. As Scharmer (2009) has emphasized, we need to put together insights from organizational learning, action inquiry, multi-stakeholder learning processes and product innovation (design thinking) to address what seems to be common sense by now, that most of society's toughest problems are not owned by any one person or institution. This means we need to carefully involve other perspectives than the usual in an ongoing RTE. In most crisis contexts, numerous humanitarian organizations are involved and jointly produce a result that cannot be attributed to the sum of their individual action. Hence, Inter-Agency approaches seem to present a valuable approach that could be applied increasingly by NGO consortia or NGO fora. Perspectives that could be included in drafting terms and informing/managing the RTE process include internal experts, planning staff, local stakeholders such as religious leaders or community authorities as well as – where possible and feasible – relevant sector experts from the affected communities or local organizations. Where their personal participation is not feasible, their perspective should be included through consultation during the early stages of an RTE and at critical points in the evaluation process.

Inflexible adherence to pre-planned Logical Framework tables and an overly rigid application of Project Cycle Management tools, on the other hand, may inhibit adaptive changes. Teams on the ground would be better equipped with a process that allowed constant and on the spot feedback by external facilitators (be they in-house or consultants).

4. Trigger Them by Demand Only

RTE needs to be demand-driven by responsible stakeholders who are willing to learn and to take action based on the findings (McNall & Foster-Fishman, 2007). Automatic triggering as foreseen for IA-RTE is thus not the right option, because it is blind to demand and potential for change. Rather, those targeted need to be *able and willing* to learn and change, i.e. there needs to be room for improvement and sufficient authority, as well as sufficient flexibility in the overall program setup. Hence, early agreement between leadership and staff on the possibility to do an RTE is essential, so it can then be realized quickly when the need arises.

5. Harness Relevant Feedback from Outside the Organization

Finally, while being an internally-controlled process, RTE should harness relevant feedback from outside their organizations, i.e. from those targeted by their action (beneficiaries) but also from outside their target group. However, requesting implementation partners to conduct beneficiary surveys can create a conflict of interest and they may chose not to feed back compromising information (FAO, 2007: 57). Again, an explicit organizational learning focus is vital to allow for the necessary openness. To facilitate efficient collection and processing of data, new technological solutions to gather beneficiary feedback have become available since the early phases of Real-time Evaluation (cf. UN Global Pulse Initiative, n.d.). Their use could facilitate quick and direct feedback from affected populations and narrow the information gap between implementers and recipients of services in real-time, allowing RTE to make on-going projects and programs more effective.

Conclusion

We have seen that the RTE approach promises to meet considerable demand in the humanitarian sector. Its application in practice, however, is more varied than standard definitions suggest. Given their seemingly simple and cost-saving nature, the risk of misappropriate use and crowding out of more rigorous evaluations is significant. The approach should thus be reserved for the narrow niche it was intended and be implemented with an explicit learning focus. Based on an analysis of current practice, we suggested ways to optimize RTE use in humanitarian assistance. Future research will have to show whether these options are reasonable in terms of allocated time and resources and whether humanitarian organizations will be able to actually use results in the overall management of humanitarian action.

References

Ackoff, R. L. (1999). *Ackoff's best: his classic writings on management*. New York, NY: Wiley.

Argyris, C. (1990). *Overcoming organizational defenses: facilitating organizational learning*: Allyn and Bacon.

Argyris, C., & Schön, D. A. (1978). *Organizational learning*. Reading, Mass.: Addison-Wesley Pub. Co.

Clark, P., & Ramalingam, B. (2008). Organizational change in the humanitarian sector. In J. Mitchell (Ed.), *ALNAP seventh review of humanitarian action*. London: ALNAP.

Collinson, S., & Elhawary, S. (2012). Humanitarian space: a review of trends and issues. London: Humanitarian Policy Group (HPG) / Overseas Development Institute (ODI).

Cosgrave, J., Ramalingam, B., & Beck, T. (2009). Real-time evaluations of humanitarian action. An ALNAP Guide: ALNAP / ODI.

Crisp, J. (2000). Thinking outside the box: evaluation and humanitarian action. *Forced Migration Review, 8*, 4-7.

Crisp, J., Martin, L., & Prattley, W. (1992). Review of UNHCR emergency preparedness and response in the Persian Gulf crisis (GULF/EVAL/12, Rev. 1). Geneva: UNHCR.

DAC (n.d.). DAC criteria for evaluating development assistance. Retrieved from http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

DG ECHO. (2007). Evaluation of humanitarian aid by and for NGOs: European Commission.

Herson, M., & Mitchell, J. (2005). Real-Time Evaluation: where does its value lie? . *Humanitarian Exchange Magazine*(32).

Hobday, M. (1988). Evaluating collaborative R&D programmes in information technology: The case of the U.K. Alvey programme. *Technovation, 8*, 271-298.

IASC (2012). Responding to level 3 emergencies: the humanitarianprogramme cycle. Retrieved from http://www.humanitarianinfo.org/iasc/downloadDoc.aspx?docID=6462

IA RTE Support Group. (2010). IA RTE procedures and methodologies. Geneva / New York: United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

Jamal, A., & Crisp, J. (2002). Real-time humanitarian evaluations. Some frequently asked questions. Geneva: United Nations High Commissioner for Refugees (UNHCR).

Jones, H. (2011). Taking responsibility for complexity. London: Overseas Development Institute (ODI).

Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy. *IBM Systems Journal*. Retrieved from http://alumni.media.mit.edu/~brooks/storybiz/kurtz.pdf

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues, 2*(4), 34-46.

Ling, T. (2012). Evaluating complex and unfolding interventions in real time. *Evaluation, 18*(1), 79-91.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*(1), 35-57.

McNall, M., & Foster-Fishman, P. G. (2007). Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation, 28*(2), 151-168.

Mitchell, M. (2009). *Complexity: A guided tour*: Oxford University Press.

Oxfam. (n.d). Real-Time Evaluation. Retrieved from http://www.alnap.org/pool/files/oxfam.pdf

Patton, M. Q. (1997). *Utilization-focused evaluation*. Thousand Oaks, London, New Delhi: Sage Publications.

Patton, M. Q. (2011). *Developmental Evaluation. Applying complexity concepts to enhance innovation and uuse*. New York, London: The Guilford Press.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*: SAGE Publications.

Ramalingam, B., Jones, H., Reba, T., & Young, J. (2008). Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. *Working Paper 285*: Overseas Development Institute (ODI).

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

Sandison, P. (2006). The utilisation of evaluations. In J. Mitchell (Ed.), *ALNAP Review of Humanitarian Action in 2005: Evaluation utilisation* (pp. 89-144). London: ALNAP.

Scharmer, O. (2009). *Theory U: Leading from the future as it emerges*. San Francisco: Berret-Koehler.

Schein, E. H. (1969). *Process consultation: its role in organization development*: Addison-Wesley Pub. Co.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report, 2*, 36-38.

Scriven, M. (2008). Meta-Evaluation revisited. *Journal of MultiDisciplinary Evaluation [Online] 6*(11), iii-viii.

Senge, P. M. (2006). *The fifth discipline: The art & practice of the learning organization*: Crown Publishing Group.

Stufflebeam, D. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. S. Scriven & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 117-141). Boston: Kluwert-Nijhoff.

Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation, 2001*(89), 7-98. doi: 10.1002/ev.3

Stufflebeam, D. (2004). The 21st-century CIPP model. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 245-266). Thousand Oaks, California: Sage Publications.

UN Global Pulse Initiative. (n.d.). United Nations Global Pulse. Retrieved from http://www.unglobalpulse.org/roadmap

UN OCHA. (2011). Inter agency real time evaluation (IA RTE) - Phase II of the humanitarian response to the earthquake in Haiti. Call for Expression of Interest. Geneva / New York: United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Retrieved from: http://www.alnap.org/pool/vacancies/eoi-ia-rte-haiti-phase-ii.pdf

van de Putte, B. (2001). Findings of the ALNAP commissioned study: Improving follow-up to evaluations of humanitarian programmes. London: ALNAP.

Weaver, W. (1948). Science and complexity. *American Scientist, 36*, 536-544.

Weick, K. E. (1995). *Sensemaking in organizations*: SAGE Publications.

Weiss, C. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis, 3*(4), 531-545.

White, H. (2009). Some reflections on current debates in impact evaluation *3ie Working Paper 1*: International Initiative for Impact Evaluation (3ie).

Zelik, D. J., Patterson, E. S., & Woods, D. D. (2010). Measuring attributes of rigor in information analysis. In E. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams*. Aldershot, UK: Ashgate.

Appendix A

Reports included in meta-evaluation

CARE International

Baker, J., & Chikoto, G. (2004). Final report of Care International's humanitarian response to the Darfur crisis. Real-Time Evaluation (RTE) phase #1. Retrieved from

http://www.alnap.org/pool/files/erd-3266-full.pdf
Baker, J. M., Hikmat, F., & Vaughan, S. G. (2004). Care International's response to the Darfur Crisis. Real-Time Evaluation (RTE) phase #2. Retrieved from http://www.alnap.org/pool/files/erd-3267-full.pdf

Catholic Relief Service (CRS)

Anonymous/CRS (2010). Real Time Evaluation of CRS' flood response in Pakistan Jacobabad and Kashmore, Sindh. Retrieved from http://www.alnap.org/resource/5991.aspx
Johnson, A. (2011). CRS Haiti Real Time Evaluation of the 2010 earthquake response: Findings, recommendations, and suggested follow up. Retrieved from http://www.alnap.org/pool/files/crs-haiti-rte-summary-and-follow-up-plan-march2011.pdf
Hagens, C. (2010). Real-Time Evaluation report for the CRS Pakistan response in the Swat Valley. Retrieved from http://www.alnap.org/resource/5821.aspx

Disasters Emergency Committee (DEC)

Murtaza, N. et al. (2011). Pakistan floods 2010. The DEC Real-Time Evaluation report. Retrieved from http://www.dec.org.uk/sites/default/files/files/Evaluations/Pakistan/Pakistan%20RTE%20-%20May%202011.pdf

ECHO

Grünewald, F., Kauffmann, D., Boyer, B., & Patinet, J. (2011). Real-time evaluation of humanitarian action supported by DG ECHO in Haiti 2009 – 2011. Retrieved 05/20/2011, from http://reliefweb.int/sites/reliefweb.int/files/resources/Groupe-URD_evaluation_Haiti_en.pdf
Hoogendoorn, M. A., & Chisvo, M. M. (2008). Real-Time Evaluation of DG ECHO Financed Action of Care International Deutschland in Zimbabwe: ECHO. Retrieved 05/20/2011, from http://ec.europa.eu/echo/files/policies/evaluation/2008/zimbabwe_RTE-Care.pdf

Food and Agriculture Organization (FAO)

FAO. (2007). Real Time Evaluation of the FAO emergency and rehabilitation operations in response to the Indian Ocean earthquake and tsunami. Rome: FAO office of evaluation. Retrieved from http://www.fao.org/docs/eims/upload/227090/RTEFinalReport.doc
FAO. (2010a). Second Real-Time Evaluation of FAO's work on highly pathogenic avian influenza (Hpai). Rome: FAO Office of Evaluation. Retrieved from http://www.fao.org/docrep/meeting/019/k8501e.pdf
FAO. (2010b). Second Real-Time Evaluation of FAO's work on highly pathogenic avian influenza (Hpai): Management response. Rome: FAO. Retrieved from http://www.fao.org/docrep/meeting/019/k8501e01.pdf
Shallon, D., & Hall, D. (2007). The first Real-Time Evaluation of FAO's work on highly pathogenic avian influenza. Rome: FAO. Retrieved from http://www.alnap.org/pool/files/918.pdf

Humanitarian Accountability Project (HAP)

Christoplos, I. (2002). The Humanitarian Accountability Project field trial in Afghanistan. Retrieved from http://www.alnap.org/resource/3095.aspx
Christoplos, I. (2002). The Humanitarian Accountability Project first trial in Sierra Leone. Retrieved from http://www.hapinternational.org/pool/files/430-SL%20July%2016%20HAP%20Evaluation%20LOGO.pdf

Inter Agency Standing Committee (IASC)

Binder, A. (2009). Real Time Evaluation of the GenCap Project. Retrieved from http://www.gppi.net/fileadmin/gppi/GPPi_final_report_GenCap_Real_Time_Eva.pdf
Broughton, B., & Maguire, S. (2006). Inter-agency Real-Time Evaluation of the humanitarian response to the Darfur crisis. New York, Geneva: IASC / OCHA. Retrieved from http://documents.wfp.org/stellent/groups/public/documents/reports/wfp092382.pdf
Broughton, B. & Maguire, S. (2006). Inter-agency Real-Time Evaluation of the humanitarian response to the Darfur crisis. Retrieved from http://documents.wfp.org/stellent/groups/public/documents/reports/wfp092382.pdf

Cosgrave, J., Polastro, R., Zafar, F. (2010). Inter-Agency Real-Time Evaluation of the humanitarian response to Pakistan's 2009-2010 displacement crisis. Retrieved from http://daraint.org/wp-content/uploads/2010/10/IARTE_PK_displacement_2010_report.pdf

Cosgrave, J., Gonçalves, C., Martyris, D., Polastro, R., & Sikumba-Dils, M. (2007). Inter-agency Real-Time Evaluation of the response to the February 2007 floods and cyclone in Mozambique: UN OCHA / IASC. Retrieved from http://ochanet.unocha.org/p/Documents/MZ_RTE_Report_Final%5B1%5D.pdf

Grünewald, F., & Binder, A. (2010). Inter-agency Real-Time Evaluation in Haiti: 3 months after the earthquake: Groupe URD / GPPi. Retrieved from http://www.urd.org/IMG/pdf/Haiti-IASC_RTE_final_report_en.pdf

Grünewald, F., Robins, K., Odicoh, A., Woldemariam, M., Nicholson, & N., Teshome, A. (2006). Real Time Evaluation of the drought response in the Horn of Africa. Retrieved from http://ochanet.unocha.org/p/Documents/RTE_Horn_Of_Africa_Regional_Synthesis.pdf

Polastro, R., Roa, B., & N. Steen (2010). Inter-agency Real Time Evaluation (IA-RTE) of the humanitarian response to Typhoons Ketsana and Parma in the Philippines. Retrieved from http://www.alnap.org/node/6949.aspx

Polastro, R., Nagrah, A., Steen, N. & Zafar, F. (2011). Inter-agency Real Time Evaluation of the humanitarian response to Pakistan's 2010 flood crisis. Retrieved 08/28/2012, from http://daraint.org/wp-content/uploads/2011/03/Final-Report-RTE-Pakistan-2011.pdf

Young, N., Gul Khattak, S., Bengali, K., & L. Elmi (2007). IASC Inter-agency Real-time Evaluation of the Pakistan floods/cyclone Yemyin. Retrieved from http://www.alnap.org/resource/3556.aspx

International Federation of the Red Cross (IFRC)

Bhattacharjee, A. (2005b). Real Time Evaluation of tsunami response in Asia and East Africa, second round. Geneva: International Federation of Red Cross and Red Crescent Societies (IFRC). Retrieved from http://www.alnap.org/pool/files/937.pdf

Herson, M. (2005a). A fine balance. Asia earthquake and tsunamis: Real Time Evaluation report for the Federation of Red Cross and Red Crescent Societies #1. Geneva: International Federation of Red Cross and Red Crescent Societies (IFRC). Retrieved from http://www.ifrc.org/Global/Publications/disasters/tsunami/RTE1synthesis.pdf

International Federation of Red Cross and Red Crescent Societies (IFRC). (2006). Real Time Evaluation (first round) - recommendations and responses. Geneva. Retrieved from www.ifrc.org/Global/Publications/disasters/tsunami/RTE1responses.pdf

International Federation of Red Cross and Red Crescent Societies (IFRC). (n.d.). Real Time Evaluation (second round) - recommendations and responses. Geneva. Retrieved from www.ifrc.org/Global/Publications/disasters/tsunami/RTE2responses.pdf

Oxfam

Manish Kumar Agrawal, M.N., Scott, I., Farhodi, W.M., Walden, V.M. (2007). Real Time Evaluation of Oxfam GB's response to the South Asia floods July-September 2007. Retrieved from http://www.alnap.org/pool/files/erd-3605-full.pdf

Simpson, R., Legesse, N. B., & Mubayiwa, R. (2009). Real Time Evaluation of the cholera response in Zimbabwe: Oxfam. Retrieved from http://www.alnap.org/pool/files/erd-3669-full.pdf

Tinnemans, K, Rowley, C., Ansari, A., & Blackwell, H. (2009). Real Time Evaluation East Asia region typhoon Ketsana/Ondoy and West Sumatra earthquake. Retrieved from http://www.alnap.org/pool/files/eastasiaresponse2009.pdf

Walden, V.M., Briere, J.F., Kumar, P. (2008). Real Time Evaluation of Oxfam International's response to cyclone Sidr November 2007-January 2008. Retrieved from http://www.alnap.org/pool/files/erd-3588-full.pdf

Tearfund

Goyder, H. (2010). Real Time Evaluation of Tearfund's Haiti earthquake response. Retrieved from http://www.alnap.org/pool/files/tearfund.pdf

UNHCR

Bartsch, D., & Belgacem, N. (2004). Real Time Evaluation of UNHCR's response to the emergency in Chad. Retrieved from http://www.alnap.org/pool/files/erd-3213-full.pdf
Bourgeois, C., Diagne, K., & Tennant, V. (2007). Real Time Evaluation of UNHCR's IDP operation in the Democratic Republic of Congo. Retrieved from http://reliefweb.int/sites/reliefweb.int/files/resources/1F5AF2A6AB229F70C125736400472B72-UNHCR-Sep2007.pdf
Bourgeois, C., Wright, N., & Crisp, J. (2007). Real-Time Evaluation of UNHCR's IDP operation in Uganda. Retrieved from http://www.unhcr.org/research/RESEARCH/46c1b8b92.pdf
Crisp, J., Graf, A., & Tennant, V. (2010). Banking on solutions. A Real-Time Evaluation of UNHCR's shelter grant programme for returning displaced people in Northern Sri Lanka. Retrieved from http://www.unhcr.org/refworld/country,,UNHCR,,LKA,,4c47fc942,0.html
Jamal, A. (2001). The Sudan/Eritrea emergency, May - July 2000. An evaluation of UNHCR's response. Retrieved from http://www.alnap.org/pool/files/erd-2924-full.pdf
Jamal, A., & Stage, O. (2001). UNHCR's programme for internally displaced people in Angola. A joint Danida/UNHCR review. Retrieved from http://www.unhcr.org/cgibin/texis/vtx/research/opendoc.pdf?tbl=RESEARCH&id=3b13a0f91
Jamal, A. & Stigter, E. (2002). Real-Time Evaluation of UNHCR's response to the Afghanistan emergency. Retrieved from http://reliefweb.int/node/102682
Savage, E., Wright, N., & Kiragu, E. (2007). Real Time Evaluation of UNHCR's IDP operation in Somalia. Retrieved from http://www.unhcr.org/cgibin/texis/vtx/search?page=search&docid=46e927652&query=Evaluation%20of%20the%20implementation%20of%20UNHCR%E2%80%99s%20policy%20on%2refugees%20in%20urban%20areas
Sperl, S., Khassim, D., & D. Snider (2006). Real-time Evaluation of UNHCR's response to the emergency in Lebanon and Syria, July - September 2006. Retrieved from http://www.alnap.org/resource/3422.aspx
Wright, N., Savage, E., & Tennant, V. (2007). Real-Time Evaluation of UNHCR's IDP operation in Liberia. Retrieved from http://www.icva.ch/doc00002439.pdf

UNICEF

Bhattacharjee, A., & Varghese, M. (2009). UNICEF's response to Georgia crisis: Real Time Evaluation. Retrieved from http://www.unicef.org/evaldatabase/files/UNICEF_RTE_FinalReport%281%29.pdf
Steets, J., & Dubai, K. (2010). Real-Time Evaluation of UNICEF's response to the Sa'ada conflict in Northern Yemen. Retrieved from http://www.gppi.net/fileadmin/media/pub/2010/Steets_2010_Yemen_Report_UNICEF.pdf

World Food Programme (WFP)

WFP (2005). Full report of the 'Real Time' Evaluation of WFP's response to the Indian Ocean tsunami. Retrieved from http://documents.wfp.org/stellent/groups/public/documents/reports/wfp079977.pdf
WFP. (2009). Full report of the Real-Time Evaluation of WFP's response to the Southern Africa crisis, 2002-2003 (Emop 10200). Retrieved from http://documents.wfp.org/stellent/groups/public/documents/reports/wfp022512.pdf

Appendix B

Table 1
List of Interviewees

Name	Organization
Andrea Binder	Global Public Policy Institute (GPPi)
Andreas Schuetz	UN OCHA
Francois Grunewald	Groupe URD
Jeff Crisp	UNHCR
John Telford	Independent
Julia Steets	Global Public Policy Institute (GPPi)
Rachel Bedouin	FAO