DESK REVIEW OF REAL-TIME EVALUATION EXPERIENCE

PETA SANDISON

For every child
Health, Education, Equality, Protection
ADVANCE HUMANITY

unicef

**EVALUATION
WORKING PAPER**

DESK REVIEW OF REAL-TIME EVALUATION EXPERIENCE

PETA SANDISON

Desk Review of Real-Time Evaluation Experience
© United Nations Children's Fund (UNICEF), New York, 2003

This is a working document. It has been prepared to facilitate the exchange of knowledge and to stimulate discussion.

The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors.

The designations in this publication do not imply an opinion on legal status of any country or territory, or of its authorities, or the delimitation of frontiers.

Acknowledgements

TABLE OF CONTENTS

ANNEXES:

ACRONYMS

ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
CO	Country Office
DAC	Development Assistance Committee
DEC	Disasters Emergency Committee
DFID	Department for International Development
EPAU	Evaluation and Policy Analysis Unit
EO	Evaluation Office
GROUPE URD	Groupe Urgence-Rehabilitation-Developpement
HAP	Humanitarian Accountability Project
IFRC	International Federation of the Red Cross
IMEP	Integrated Monitoring, Evaluation Planning Process
M&E	Monitoring and Evaluation
NGO	Non-Governmental Organisation
OIA	Office of Internal Audit
RO	Regional Office
RTE	Real-time evaluation
TOR	Terms of Reference
UNHCR	United Nations High Commission for Refugees
UNICEF	United Nations Children's Fund
WFP	World Food Programme

EXECUTIVE SUMMARY

UNICEF commissioned this desk study with the objective of identifying key lessons on real-time evaluation (RTE) through drawing from the experience of other agencies and organisations in this area. The findings are intended to provide the basis for UNICEF to develop and test a real-time review or evaluation methodology as a component of its evaluation system. A number of real-time evaluations were examined from seven humanitarian agencies and interviews carried out with more than 20 individuals with experience of RTE.

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. The common characteristics of RTE examined in this study are that they correspond to standard definitions and characteristics of evaluation, are carried out in the early stages of an emergency, typically two to three months after the crisis and ideally, though not necessarily, repeated during the project cycle. The evaluators are either internal but not directly involved with or responsible for the programme in question, or external but highly familiar with the agency's work. The approach emphasises participation with agency staff and the reporting method prioritises accessibility, particularly rapid dissemination and participation with the implementing staff. Hence, findings and recommendations are delivered briefly in verbal and written form, typically prior to leaving the field, and the length of even final reports is kept short.

Agencies interviewed saw real additional value gained by using RTE as compared to internal review or standard evaluations, mainly because of its timing and rapid feedback, and its combination of proximity to the actual programme and people with its evaluative distance and objectivity. Its recommendations can be checked for appropriateness in the field prior to departure and fed straight back into the programme, making an immediate difference and enhancing accountability. At the same time, its contact with staff early in an emergency, its witness to policy in practice and privileged access to process and other unrecorded information can capture unique information that enhances subsequent standard evaluations and, potentially, organisation-wide learning. It facilitates *in situ* learning that may be a highly appropriate form of learning for field staff. The participative process of RTE can facilitate team building and resolve tensions by bridging the gap between the country office, regional office and headquarters, bringing up-to-date information from one to another and acting as "a voice" or advocate when appropriate. An RTE team can also be a resource encouraging and advising on appropriate baseline surveys and monitoring systems.

Many of the shortcomings of RTE are associated with it being done badly rather than with it being done at all. Nonetheless, it has inherent risks primarily associated with its timing. Although its recommendations may save money, it is likely to increase the evaluation budget of a programme and may be difficult to fund and to recruit for, particularly as a high calibre of evaluators is emphasised and there is a short time period in which to implement, after which the opportunity to be genuinely real-time is lost. An RTE may over-emphasise what is essentially a snapshot of a fast-moving situation; the picture lending too much weight to ephemera and becoming hard to subsequently shake off. Negative, published reports could damage the agency's reputation and relationships, undermining its ability to function during an emergency. Programme staff may be too pressurised to absorb lessons and the evaluation team could distract them from life-saving activities. A greater reliance than usual on interviews and the need for the evaluator(s) to establish good relationships and act as a resource within a team could distort the evaluator's role of objectivity and distance from management.

The benefits of an RTE are increased through a full process of consultation with key (internal) stakeholders in developing the TOR, and clarity and careful planning prior to implementation regarding its purpose, for example the extent to which it should serve external or exclusively internal accountability and hence the role of the evaluator(s), types of reporting and mechanisms for follow-up of recommendations. The process of the RTE is important in terms of maximising participation and ownership of the findings and recommendations and enabling rapid feedback and implementation.

RESUME

L'UNICEF a lancé cette étude théorique afin d'arriver à connaître les leçons qu'on pouvait tirer de l'évaluation en temps réel (ETR) à partir de l'expérience acquise en la matière par d'autres agences et organisations. Les résultats obtenus doivent fournir une base permettant à l'UNICEF de développer et de tester une procédure d'examen en temps réel ou une méthode d'évaluation venant s'intégrer dans son système d'évaluation. Un certain nombre d'évaluations en temps réel ont été étudiées dans sept organisations humanitaires et des entretiens ont été réalisés avec plus de 20 personnes disposant d'une expérience de l'ETR.

On effectue une évaluation en temps réel lorsqu'un programme est pleinement opérationnel et les conclusions de l'ETR sont retransmises presque simultanément au programme, pour une prise en compte immédiate. Les ETR analysées dans cette étude ont en commun certaines caractéristiques: elles correspondent à des définitions et des caractéristiques d'évaluation classiques, elles sont menées dans les premiers temps d'une situation d'urgence, généralement dans les deux ou trois mois qui suivent la crise, et il est souhaitable, sans que cela soit une nécessité, qu'on procède à d'autres ETR durant le cycle du projet. Ceux qui procèdent à l'évaluation peuvent appartenir à l'organisation, sans être directement impliqués dans le programme en question ou en être responsables, ou bien venir de l'extérieur, tout en ayant une grande connaissance du travail qu'effectue l'organisation. Dans cette démarche, on met l'accent sur une implication du personnel de l'organisation et la méthode de compte rendu insiste sur l'accessibilité, la très grande rapidité de diffusion et la participation du personnel chargé de la mise en oeuvre. Aussi les conclusions et recommandations sont-elles communiquées brièvement, verbalement et par écrit, en général avant de quitter le terrain, et on veille à ce que les rapports définitifs eux-mêmes soient courts.

Les organisations interrogées ont estimé que l'ETR offrait une valeur ajoutée par rapport à l'examen interne ou aux évaluations classiques, surtout en raison de son calendrier d'intervention et d'une remontée de l'information rapide, et du fait qu'elle alliait d'une part une proximité vis-à-vis du programme lui-même et des personnes impliquées et d'autre part un recul et une objectivité pour l'évaluation. Le bien-fondé des recommandations de l'ETR peut être examiné sur le terrain avant que ces recommandations ne remontent et ne reviennent sur le programme, ce qui fait avancer les choses et responsabilise. En outre, l'ETR permet un contact avec l'équipe sur le terrain, dès le début de la situation d'urgence ; elle constitue un témoignage de la politique telle qu'elle se traduit dans les faits et elle implique un accès privilégié à l'opération et à d'autres informations non enregistrées, ce qui peut permettre d'acquérir des connaissances exceptionnelles, qui faciliteront les évaluations classiques venant par la suite et, éventuellement, l'acquisition du savoir à l'échelle de l'organisation. L'ETR favorise l'apprentissage « sur le tas », ce qui peut constituer une forme particulièrement appropriée d'acquisition des connaissances pour le personnel sur le terrain. La procédure de participation qu'on trouve dans l'ETR peut faciliter la formation d'équipes et régler des tensions en comblant le fossé existant entre le bureau dans le pays, le bureau régional et le siège, grâce à la transmission de l'un à l'autre d'informations à jour et au rôle de porte-parole ou d'avocat que joue l'ETR lorsqu'il le faut. Une équipe d'ETR peut être également une source d'encouragements et de conseils pour les enquêtes de base adaptées et les systèmes de suivi.

Dans un bon nombre de cas, les défauts de l'ETR surviennent lorsqu'elle est mal faite et ils ne correspondent pas à la nature de l'ETR. Néanmoins, il existe des risques inhérents à l'ETR, essentiellement liés à son calendrier. Bien que ses recommandations permettent des économies, l'ETR va probablement augmenter le budget d'évaluation d'un programme ; en outre, il peut s'avérer difficile de financer l'ETR et de recruter pour la réaliser, en particulier du fait qu'on insiste sur la nécessité d'évaluateurs de haut niveau et qu'on dispose de peu de temps pour mettre l'opération en place et saisir l'opportunité d'agir vraiment en temps réel. Une ETR peut insister exagérément sur ce qui est essentiellement un instantané dans une situation en évolution rapide ; la photo accorde trop d'importance à l'éphémère et il est difficile de s'en débarrasser par la suite. La publication de rapports négatifs peut nuire à la réputation de l'organisation et la gêner dans ses relations, réduisant ainsi ses capacités de fonctionner en situation d'urgence. L'équipe du programme peut être soumise à des pressions excessives pour assimiler les leçons données et les évaluateurs pourraient la gêner dans ses actions de sauvetage. Un recours exagéré aux entretiens, le besoin ressenti par le ou les évaluateurs d'établir de bonnes relations et d'agir comme un expert au sein de

l'équipe pourrait rendre difficile l'objectivité de l'évaluateur et le respect des distances vis-à-vis de l'équipe gérant l'opération.

Les avantages d'une ETR augmentent grâce à un processus complet de consultation des parties prenantes clés (internes), en développant les termes de référence, une planification dans la clarté et approfondie avant la mise en oeuvre par rapport à son objectif. Par exemple, dans quelle mesure doit-elle servir la responsabilité externe ou la responsabilité exclusivement interne et donc quel doit être le rôle du ou des évaluateurs, les types de compte rendu et les mécanismes de suivi des recommandations? Le processus de l'ETR est important dans la mesure où il favorise l'amélioration de la participation et la prise à son compte des conclusions et recommandations, et permet une remontée d'information et une mise en place rapides.

RESUMEN EJECUTIVO

El UNICEF encomendó el presente estudio de gabinete con el objetivo de determinar, sobre las base de la experiencia de otros organismos y organizaciones, cuáles son las lecciones fundamentales que arroja la evaluación en tiempo real (RTE). El propósito de las constataciones de la RTE es proporcionar una base para que el UNICEF prepare y ponga a prueba una metodología de examen o evaluación en tiempo real, como componente de su sistema de evaluación. Se examinaron varias evaluaciones en tiempo real efectuadas por siete organismos de asistencia humanitaria y se realizaron entrevistas con más de 20 personas poseedoras de experiencia en la RTE.

La evaluación en tiempo real se realiza cuando un programa está en pleno curso de ejecución; y la retroinformación sobre las constataciones de la evaluación se aporta al programa casi simultáneamente, para su inmediata utilización. Los casos de RTE examinados en este estudio tienen características comunes: se ajustan a definiciones y características estándar de la evaluación, se realizan en las etapas iniciales de una emergencia (típicamente dos o tres meses después del comienzo de la crisis) y además, idealmente, aun cuando no necesariamente, se repiten durante el ciclo del proyecto. Los evaluadores son o bien internos (pero no participantes en el programa en cuestión ni directamente responsables de éste), o bien externos (pero sumamente familiarizados con la labor del organismo). Esta modalidad hace hincapié en la participación junto con los funcionarios del organismo; asimismo, en el método de presentación de informes se asigna prioridad a la accesibilidad, en particular la rápida difusión y el involucramiento del personal encargado de la ejecución. Por consiguiente, las constataciones y recomendaciones de la RTE se presentan sucintamente, en forma oral y escrita, mayormente antes de que los evaluadores se marchen, y los informes, incluso los informes finales, son breves.

En las entrevistas con organismos se puso de manifiesto que, a juicio de éstos, la RTE agrega valor, en comparación con los exámenes internos o evaluaciones externas, debido principalmente al momento en que se efectúa, a la rápida retroinformación y a la combinación de su cercanía al programa y los funcionarios, con el distanciamiento y la objetividad de la evaluación. Las recomendaciones de la RTE pueden verificarse sobre el terreno antes de la partida de los evaluadores y pueden incorporarse en el programa, de modo de mejorarlo de inmediato y fortalecer la rendición de cuentas. Al mismo tiempo, el contacto de los evaluadores con el personal poco tiempo después de iniciarse la situación de emergencia, su percepción directa de las políticas tal como se las aplica y su acceso privilegiado a los procesos y a otros datos aún no registrados, puede captar una valiosísima información que ha de mejorar las ulteriores evaluaciones estándar y, potencialmente, el aprendizaje en toda la organización. La RTE facilita el aprendizaje *in situ*, el cual puede ser sumamente apropiado para el personal que trabaja sobre el terreno. Los procedimientos participativos de la RTE pueden facilitar la constitución de equipos y resolver tensiones, salvando las distancias entre la oficina en el país, la oficina regional y la sede, aportando información actualizada de las unas a las otras y actuando como "portavoz" o "factor de promoción", según proceda. Un equipo de RTE también puede ser un recurso que aliente y asesore la realización de encuestas apropiadas para obtener información básica de referencia y establecer sistemas adecuados de monitoreo.

Entre los defectos que se han señalado en la RTE, muchos pueden atribuirse a defectos en su realización y no al hecho de que se realice. Con todo, la RTE conlleva riesgos intrínsecos debido principalmente al momento en que se realiza. Aun cuando sus recomendaciones pueden redundar en economías, lo probable es que la RTE acreciente el presupuesto de evaluación en un programa y que sea difícil financiarla y contratar personal, particularmente debido a la enorme importancia de que los evaluadores sean de alto calibre profesional y a la brevedad del lapso disponible para llevarla a cabo, pasado el cual se pierde la oportunidad de efectuar una auténtica RTE. Es posible que en una RTE se exagere la importancia de lo que es esencialmente una instantánea de una situación en acelerada evolución; y que en esta instantánea se asigne demasiada importancia a aspectos efímeros y que ulteriormente sea difícil descartarla. Al publicar informes negativos, es posible dañar la reputación del organismo y sus relaciones, menoscabando así su capacidad operativa durante una emergencia. También es posible que el personal de programas esté sujeto a presiones de gran magnitud y no pueda absorber las lecciones; y también es posible que el equipo de evaluación lo distraiga de actividades que salvan vidas. Asimismo, es posible que un grado de dependencia superior a lo habitual respecto de las entrevistas, sumado a la necesidad de que los evaluadores

establezcan buenas relaciones y actúen como fuentes de información dentro de un equipo, pudiera distorsionar la objetividad del papel del evaluador y su distanciamiento de la administración.

Los beneficios de la RTE se acrecientan cuando media un cabal proceso de consultas con los interesados directos (internos) clave en la preparación del mandato (términos de referencia, TOR), así como una clara y cuidadosa planificación antes de la puesta en práctica, en lo relativo a su propósito, por ejemplo, determinar en qué medida debe estar al servicio de una rendición de cuentas externa, o exclusivamente interna, y por ende, cuáles han de ser el papel de los evaluadores, los tipos de informes a presentar y los mecanismos de seguimiento y formulación de recomendaciones. El proceso de la RTE es importante en lo concerniente a maximizar la participación y la apropiación de las constataciones y recomendaciones, además de posibilitar una rápida retroinformación y una puesta en práctica sin tardanza.

DESK REVIEW OF REAL-TIME EVALUATION EXPERIENCE

1.   BACKGROUND

1.1   Background to This Review

As part of an organisation-wide process undertaken by the United Nations Children's Fund (UNICEF) to strengthen its humanitarian response capacity, the organisation's Evaluation Office (EO) in collaboration with the Office of Internal Audit (OIA) is pursuing a number of strategies to strengthen monitoring and evaluation (M&E) at field level. One of these strategies is to establish some form of real-time evaluation (RTE). The purpose of this desk study is to identify key lessons on real-time evaluation, drawing from the experience of other agencies and organisations in this area. The findings are intended to provide the basis for UNICEF to develop and test a real-time review or evaluation methodology as a component of its evaluation system. Concerns have been expressed within UNICEF regarding anticipated reactions of defensiveness and distrust at field level if a new evaluation mechanism were to be introduced, particularly for the immediate post-crisis phase of humanitarian response as intended. Although UNICEF will seek a less threatening name for the mechanism as it is developed, this desk review will refer to real-time evaluation, the term common to a number of agencies and organisations advancing work in this area. This review is confined to humanitarian response and all descriptions and analysis refer to the use of RTE in this context. The Terms of Reference (TOR) is attached in Annexe I.

1.2   A Pocket History of Real-time Evaluation

How RTE first made its way into humanitarian policy and practice is unknown[1]. The idea of carrying out an evaluation during implementation was recommended in Hallam's Good Practice Review in 1998[2]. The earliest humanitarian RTE in practice appears to be an unpublished one commissioned by Danida in 1999 of the Kosovo emergency[3]; it does not seem to have contributed significantly to RTE gaining currency in the humanitarian world[4]. The chief proponent and implementer of RTE for humanitarian evaluation has been the United Nations High Commission for Refugees (UNHCR) Evaluation and Policy Analysis Unit (EPAU). UNHCR initially proposed something akin to RTE in 1992 in an evaluation of its Persian Gulf crisis[5]. Following the 1999 Kosovo emergency, a UNHCR Plan of Action in 2000 recommended the introduction of RTE. To date UNHCR has carried out four RTEs in Sudan-Eritrea (2000), Angola (2000), Pakistan (2001) and Afghanistan-Iran (2002), piloting and developing the approach over a period of three years. In 2002 EPAU published a paper responding to frequently asked questions on RTE[6]. WFP has also considerably developed RTE, following Broughton's study. The second of a planned series of three RTE missions to its Southern Africa programme has just been completed. In 2000, The Department for International Development (DFID) with the International Federation of the Red Cross (IFRC) carried out a real-time evaluation of the IFRC's India flood response. The IFRC recently completed an RTE of its Southern Africa programme. The Humanitarian Accountability Project (HAP) has carried out two RTEs of its field trials in Sierra Leone (2001 and 2002) and Afghanistan (2002).

No other RTEs have been identified, at least not by that name. The Disasters Emergency Committee (DEC) has carried out an extremely similar approach that could be an RTE in anything but name. Groupe Urgence-Rehabilitation-Developpement's (Groupe URD) Iterative Evaluation Process with Mini-Seminar Combination shares many common characteristics with RTE as does World Vision's synthesis of RTE and Action Review recently adopted by the agency. These latter agencies' approaches are therefore referred to in the comparative study below.

---

[1] Bernard Broughton, commissioned by the World Food Programme (WFP) to produce a concept paper and proposed methodology for RTE notes that RTE has long been in common use in science and technology, particularly computing.

[2] Hallam: "Even where conditions may not be ideal for an early evaluation, there is a lot to be said for a reasonably 'quick and dirty' evaluation to take place, while programmes can still be changed…Field workers may also feel more willing to engage with the evaluation if they feel that the benefits are more immediate." p46

[3] ETC, 1999.

[4] The report is not referred to in any studies or reports of RTE identified in this review.

[5] Crisp et al., *Review of UNHCR emergency preparedness and response in the Persian Gulf crisis* (GULF/EVAL/12, Rev.1) UNHCR Geneva, March 1992 quoted in Jamal A., & J. Crisp, *Real-time humanitarian evaluations. Some frequently asked questions*. UNHCR 2002.

[6] Jamal et al., *Real-time humanitarian evaluations. Some frequently asked questions*.

1.3   Methodology

What little "literature" exists on RTE essentially consists of Jamal and Crisp's paper on frequently asked questions for UNHCR and Broughton's concept paper for WFP. These, the RTE TORs, and reports from UNHCR, WFP, IFRC, HAP, DFID, ETC and the DEC have been used in combination with semi-structured interviews to examine the rationale for RTE, its purpose, design, advantages and disadvantages, and where possible to identify examples of good practice, challenges, and pitfalls.

Interviews were carried out mostly by telephone with 20 individuals with experience conducting RTE. Interviews by telephone and e-mail have also been carried out with five programme staff from three different agencies who have participated in RTE.

1.4   Is Real-Time Evaluation Unique?

What is RTE and why should UNICEF consider adding it to its portfolio of programme monitoring and evaluation mechanisms? Is it a new name for an old procedure, synonymous with monitoring or review, a synthesis of existing practices or an exciting innovation that can enrich the learning and accountability potential of humanitarian agencies?

UNHCR describes RTE as "a timely, rapid and iterative peer review of a fast evolving humanitarian operation (usually an emergency) undertaken at an early phase. Its broad objective is to gauge the effectiveness and impact of a given UNHCR response, and to ensure that its findings are used as an immediate catalyst for organisational change"[7].

Broughton proposes the definition "A real-time evaluation of a WFP operation is a rapid assessment conducted during implementation, often at more than one stage (including the initial stage), of the operation's relevance and design, progress in achieving the operation's objectives (i.e. results), any gaps or unintended impact, the effectiveness and efficiency of the mode of implementation, and the appropriateness and application of operational guidelines and policies"[8].

The common ground is that an RTE is rapid and examines a programme against recognisable evaluation criteria whilst it is being implemented.

The common characteristics of RTE examined in this study are that they:
- correspond to standard definitions and characteristics of evaluation;
- are typically carried out in the early stages of an emergency and, ideally though not necessarily, repeated during the project cycle;
- are carried out by teams or individuals who are not directly involved with or responsible for the programme in question;
- ensure that the reporting method prioritises accessibility, particularly rapid dissemination and participation with the implementing staff. Hence findings and recommendations are delivered briefly in verbal and written form, typically prior to leaving the field, and the length of final reports is short.

To avoid adding to the terminology of learning and accountability without adding anything to its toolbox, it is important that RTE should defend or moderate its title depending on whether it is distinguishable from other, established mechanisms such as review and monitoring[9].

Like monitoring and review, RTE takes place during a programme with the intention of making in situ changes. Unlike them, RTE uses standard evaluation criteria and methodology[10] and staff who are not the implementers of the programme. This study considers that RTE shares many of the qualities of monitoring

---

[7] Ibid.
[8] Broughton.
[9] This is discussed in more detail in Annexe I.
[10] For example, ALNAP's definition and characteristics such as: systematic, impartial, intended to support lesson-learning and accountability, utilising recognisable evaluation criteria and providing outputs in the form of written findings and recommendations. (p201 ALNAP Review 2002)

2

and review. It has great potential to link them with a standard evaluation. However, it is nonetheless an evaluation mechanism as it utilises an evaluation framework, is intermittent rather than continuous and the evaluators are not the implementers of the programme or directly responsible for its planning and decision-making.

The defining characteristic of an RTE *vis-à-vis* a standard evaluation is <u>the timing</u>, which enables it to "act as an immediate catalyst for organisational and operational change"[11]. It is the timing that creates variations from standard evaluations of purpose and objectives, as well as the usual (institutional) learning and accountability objectives of evaluation.[12] Rapid feedback is as important as the timing of the evaluation mission and without it the evaluation is not real-time[13].

Exactly what an RTE is for and how to do one may vary across the spectrum of evaluation purpose and methodology in the same way as a standard evaluation, but the term itself defines the essential characteristics of real-time evaluation, which could be summarised as:

> *a rapid evaluation carried out during the early stages of a response and possibly at intervals throughout the response in accordance with recognised evaluation criteria. It is carried out by individuals external to the organisation or internal but with no managerial or implementing responsibilities for the programme and its findings are immediately disseminated.*

2.   A COMPARATIVE ANALYSIS OF REAL-TIME EVALUATION IN THE HUMANITARIAN SECTOR

2.1   Why Use RTE?

2.1.1   Purpose

Most agencies expect the RTE to simultaneously benefit the actual programme as well as broader, institutional learning[14]. For example, WFP's pilot in Southern Africa places higher value on the iterative potential of RTE to deepen understanding and organisational learning[15], while nonetheless designing the TOR to support immediate feedback and potential change to the programmes in hand.

All those interviewed felt that RTE was addressing many of the shortcomings of standard evaluations and brought additional advantages, and they planned to continue using it.

2.1.2   Immediate Benefits

That an RTE can feed back recommendations for action "in time to make a difference"[16] to the programme being evaluated is an obvious advantage. In addition, interviewees and reports noted added value in the type of information gathered and in the real-time evaluation process itself.

Perspective
Several agencies note the value of 'outsiders', free of operational responsibilities, who are able to bring different perspectives and a critical distance into the heat of an emergency context when typically the tyranny

---

[11] Jamal RTE, Some frequently asked questions.
[12] Groupe URD's Iterative Evaluation with Mini Seminars shares many of an RTE's characteristics such as the deployment of the evaluation team during the project, repeated visits of the same evaluation team and regular communication throughout the programme's duration and the production of an Aide Memoire accompanied by a feedback workshop whilst still in the field. Where it differs is in its emphasis on longer-term interagency lesson-learning across all sectors. An iterative approach is considered essential, facilitating the accumulation and dissemination of knowledge gained on each visit. As a result the reports are considerably longer than RTE reports and produced some time after the missions.
[13] Real-time in computing is a system in which "input data is processed within milliseconds so that it is <u>available virtually immediately as feedback </u>to the process from which it is coming": The New Oxford Dictionary of English. 1998. [My underlining.]
[14] Examples of RTE purpose drawn from TORs of several agencies are included in Annexe II.
[15] WFP's Southern Africa TOR states that real-time evaluative inputs "might be the basis for corrective action…[but] …The ultimate benefit of applying RTE to WFP's emergency operations is to reach summative conclusions. "
[16] Jamal RTE, frequently asked questions.

of daily service delivery and the reactive style generated by emergencies favours the "urgent over the important", pushing strategic approaches such as advocacy and process issues to one side[17].

Pause to Reflect
Country programme staff particularly appreciated the facilitation of reflection through the discussion groups or workshops of the RTE — the "luxury" of a structured pause to permit the team itself "to assess more objectively the progress achieved"[18]. A typical review could also be expected to provide additional perspective but is more at risk of being displaced by the day-to-day exigencies and responsibilities during the early stages of an emergency.

Bridging Gaps and Managing Misconceptions
Evaluators pointed out that their presence as impartial observers helped identify the hidden constraints that may be affecting performance, such as conflict within the team and mismanaged cultural differences, and on two occasions helped to surface and resolve disagreements. The RTE report had been used as a neutral reference point by team members in order to facilitate agreement. In some cases the evaluator's up-to-date knowledge of the thinking at each level of the organisation meant that s/he was able to bridge the gap between the field and its head or regional office, acting as an advocate or mediator for the different stakeholders.

Team Learning and Morale
Having an experienced evaluation team able to bring additional knowledge to the field team of, for example, wider policy issues, current institutional thinking and lessons learned from earlier responses was often appreciated by field teams as a valuable resource that promoted personal development and learning. Several of the field participants in RTE noted a boost to team morale.

It has been suggested that the Humanitarian Sector contains a high proportion of people who learn best when "there is an obvious link between the subject matter and a problem or opportunity on the job…and have the chance to try out techniques with coaching/feedback from a credible expert"[19]. Should this be the case, the on-the-job style of learning offered by the process of evaluation in real-time could be a far more appropriate way to facilitate learning than written materials and a remote or classroom approach.

2.1.3   Long-Term Benefits

Just as an RTE can influence immediate decisions and dynamics, it can also lead to better quality standard evaluations and also potentially to improved institutional learning, though it may be too early to judge the latter. There is less evidence of changes to institutional practice through RTE organisation-wide recommendations. This may be because embedded organisational weaknesses take longer to change and most RTEs are very recent, hence it is too early to judge[20].

There are good reasons to expect that RTE could contribute to organisational learning. Evaluators commented that RTE can incorporate process and off-the-record data that would not commonly be available to standard evaluations, but that could have had a significant impact on effectiveness. The interface between policy and practice is scrutinised in an RTE, and findings can be invaluable for the development of policy and procedure. Additionally, lost information due to high staff turnover and poor record-keeping in the early stages of emergencies are the bane of standard evaluations; an RTE can mitigate such losses through its early presence. An RTE can also capture transitory events and operating contexts, particularly if it is part of a

---

[17] IFRC noted that the perspective brought by the evaluation team in Southern Africa enabled them to flag up important objectives such as sustainability and capacity building with National Societies that could otherwise have been lost in the rush.
[18] UNHCR field staff.
[19] The Active Learning Network for Accountability and Performance in Humanitarian Action's (ALNAP) Annual Review 2002 p27, quoting Honey, P. 1994 descriptions of Activist and Pragmatists learning preferences from "Styles of Learning" in Mumford, A. (ed) 1995 Gower Handbook of Management Development. Fourth Edition. Aldershot: Gower Publishing Ltd.
[20] HAP RTE recommendations are not tracked from the first to the second RTE; subsequent field trials would indicate learning. WFP and IFRC real-time evaluations are either in process or only recently completed. UNHCR has noted wider organisational changes.

series culminating in a standard evaluation by which time the fluidity of some emergencies may mean that even the location of the response has changed, sometimes more than once[21].

Finally, RTE can support the monitoring function in a programme. Many standard evaluations are unable to assess the impact of a programme because of the absence of baseline data and poor programme monitoring systems. The evaluators can facilitate the establishment of both baseline data collection and of monitoring systems early on in a programme. WFP's RTE, for example, recommended the introduction of a sentinel site method of monitoring for its Southern African response and the agency subsequently delegated its design and establishment to other consultants.

Risks of Proximity
A danger of the real-time evaluator's proximity to events is that s/he can become too close to the prevailing beliefs of the humanitarian community and replicate the same misconceptions or shortcomings[22].

2.1.4   Accountability

RTE *can* enhance accountability; how and for whom varied considerably.

Accountable to Whom?
All the organisations studied facilitated upwards, internal accountability by making their RTE findings rapidly available to, at the very least, those directly involved with the programme and senior managers and the Board of the organisation. In some cases, reports were also disseminated to a limited number of key external stakeholders.

External accountability does not appear to have been enhanced by RTE; the extent to which an RTE reaches those outside the organisation appears to be a function of the organisation's general approach to evaluation and transparency. Agencies that routinely publish (standard) evaluation reports, such as UNHCR and HAP, publish their RTE reports[23]. Unpublished DEC monitoring reports are referred to and followed up in final evaluations, which are publicly available[24].

No examples were found of greater downwards accountability through the findings or recommendations of an RTE being made directly available to the beneficiaries.

Accountable Evaluators
It could be said that the evaluators themselves are made more accountable. Their recommendations will themselves be implemented and therefore tested during the life of the programme. A final evaluation would also be evaluating the extent to which programme changes brought about by the RTE were appropriate, realistic, and relevant.

Learning Aloud
Some field staff apparently viewed an RTE with some trepidation, expecting it to be an additional, headquarters-imposed performance appraisal being carried out at a particularly pressurised moment and, potentially, leading to exposure and blame as a result. However, there is no evidence of an RTE being used as a tool for individual appraisal or censure or of findings being imposed on the teams. Almost all the agencies emphasised that it was very important to facilitate the evaluative process, not impose it and risk generating conflict, reducing ownership and therefore utility of the findings, or damaging team morale as a result. Some evaluators noted that good practice included acknowledging that many of the findings had already been identified by the staff and were in the process of being changed. Neither WFP nor UNHCR

---

[21] The DEC Goma monitoring report, for example, vividly captures the unpredictability of the environment in which agencies were making decisions and planning responses following the volcano eruption.

[22] The ETC evaluation of the Kosovo emergency for example recommends that agencies plan for a two year response. Being unable to predict a different course of action is not a criticism of the evaluation, rather a graphic example that a real-time evaluation will also lack the benefit of distance and hindsight and that users of a one-off RTE should beware of inadvertently adding weight to ephemera.

[23] WFP will publish the final RTE report following its consideration by WFP's Executive Board. WFP regards the interim RTE reports as for internal management purposes and has no plans to publish them

[24] The DEC posts some its monitoring reports on its website; each one states that it represents the views of the evaluator, not the DEC.

include formal recommendations in RTE reports preferring the use of "Action Points" to reflect consensual decision making and joint recognition of findings[25].

Open Reporting
UNHCR notes that critics of RTE emphasise "the risks of placing controversial findings in the public domain at a time when UNHCR is engaged in delicate negotiations"[26]. The publication of UNHCR's Pakistan RTE generated considerable concern in the government of Pakistan, taken by surprise at the culture of public assessment. Similarly negative findings could mark the programme for the duration, regardless of improvements. However, no agency reported this as a problem[27].

UNHCR argues that its credibility with donors and other external stakeholders has improved as a result of its policy of posting reports on its web-site. The "accounting" also reveals the constraints under which the programme is operating, as well as its successes. Staff commented that it can act as a lobbying tool; findings that point to changes beyond the control of agency staff can be clearly directed at those responsible and act as a form of advocacy.

Although it is not yet common currency, RTE could become part and parcel of the expected accountability mechanisms of an organisation. Donors may not yet request an RTE, but they probably will do.

Damage Limitation
Some agencies have reduced many of these risks through producing two-tiered reports, one internal and one external. External reports are not necessarily stripped of grit and authenticity, but use common sense to fit the report to the audience. For example, individual performance issues are properly the domain of management; the effects are still reported, but the causes are stripped out or made non-attributable. Politically sensitive or confidential issues that could damage the agencies' relationships with government, donors and partners are removed. All else can be reported.

2.2   The Design

The design characteristics particular to real-time evaluation are a function of its chosen purpose, timing and dissemination. In other respects, the scope of an RTE is essentially the same as a standard evaluation: emergency preparedness, exit strategies, capacity building, logistics, human resources, and so on.

Because RTE is working in a dynamic situation, the TORs can be opportunist by responding to emerging issues and the particular stage of the programme.

2.2.1   The Scope

Criteria
The RTE TORs available mostly use standard evaluation criteria (i.e. Development Assistance Committee [DAC] criteria) applied to and combined with agency-specific policy issues. UNHCR designs the RTE around its policy and mandate and does not use standard/DAC criteria; this is also the case with UNHCR standard evaluations. The HAP applies DAC criteria such as effectiveness to the HAP's specific objectives; similarly the IFRC uses standard evaluation criteria to examine specific programme objectives, as well as institutional goals such as capacity building. The DEC uses six criteria (largely DAC) combined with the Code of Conduct, People in Aid and Sphere as the basis of its monitoring missions. WFP has designed its RTE around the project cycle, examining each stage and all aspects of its operation through the lens of standard evaluation criteria.

---

[25] This is not to suggest that the evaluations do not feed into WFP's accountability mechanisms; the final summary report subsequently written will include formal recommendations for consideration by WFP's Executive Board.
[26] Jamal, RTE Frequently Asked Questions.
[27] Presumably this eventuality could be avoided by follow up RTE missions that also published findings.

6

Relating the Design to the Timing
The particular focus of an RTE is determined largely by its timing and the time available. In general, a real-time evaluation is more likely to encompass strategic, process and performance approaches to evaluation, rather than impact evaluation, simply by virtue of its timing. Where an RTE is based on only one field mission early in the emergency, it will focus on issues related to the early part of the project cycle, such as the appropriateness of the design, the establishment of systems and procedures, the assumptions made at the time, the quality of the assessments, etc., rather than the impact of the programme. RTEs that plan on several field trips, such as those carried out by WFP and the HAP, can design the TORs around the project cycle. WFP for example has produced a generic TOR for its real-time evaluations that is designed around a three-stage process. The first visit concentrates on the programme's needs assessment, programme design etc, the second "gradually shifts towards a focus on accountability, effectiveness and results" and the third mission addresses "findings and lessons to be learned about the intervention's overall effectiveness at achieving its objectives"[28].

Repeat missions of an RTE are generally thought to be preferable, building up a series of snap-shots, increasing understanding, following each stage of the programme and reducing the risk of evaluating an unrepresentative moment in a programme. At the same time, it appears to be something of a luxury and of the 13 RTEs examined, only four included repeat visits.

Geographic Focus
The speed with which real-time findings must be fed back at country, regional and head office levels most forcefully dictates any trade-offs between scope and speed. Logically, the tension between the two will increase in line with geographical coverage; if the field trips are drawn out, the reporting follows. UNHCR used its Afghanistan RTE visits to address both single-country and cross-border issues, by visiting first Pakistan, then Iran and Afghanistan on a second trip. This approach kept the mission's focus tight and rapid, but traded the potential benefits of repeat visits to the same country with the benefits of enhancing the organisation's regional lesson-learning and planning. Several of the RTEs examined have been carried out over large distances with very limited time, forcing evaluators to concentrate on towns and capitals. This has been particularly true of the Southern Africa RTEs[29]. There is no evidence to suggest that the broad scope of the multi-country RTEs have sacrificed quality for quantity[30], however common sense suggests that there is a risk that such RTEs could become too superficial to be useful, or conversely too unwieldy to produce the rapid feedback loop that must characterise a real-time evaluation.

For the same reasons of the time needed to both gather and process information, a system-wide evaluation is beyond the scope of RTE[31]. Additionally, on early RTE missions, the time available, particularly for external evaluators, to read secondary documentation prior to deployment may be reduced. Evaluators will also have less access in general to evaluations and reports from other agencies[32].

Selecting a Programme for an RTE
While the HAP and the DEC automatically plan an RTE as a component of their monitoring and evaluation system, agencies usually select programmes for a real-time evaluation because they are of extraordinary significance or potential impact, for example in terms of size or political importance (Southern Africa, Kosovo, Afghanistan) or controversial in some way (UNHCR in Angola for example). RTE has been most used in rapid-onset emergencies reportedly lending itself well to such contexts in which little time is available for programmes to recover from mistakes or benefit from reflection. However, it has also been used with apparently equal value in rapid-onset, slow and chronic emergencies[33]. One evaluator commented that for

---

[28] WFP TOR Southern Africa.

[29] WFP's second RTE covered six countries in about four weeks, IFRC's team covered five countries in less than three weeks, the DEC covered three countries in two weeks. The HAP methodology prioritises greater stakeholder analysis; the second field trips in both Trials were for at least two weeks, for one country. UNHCR spent two weeks in Pakistan and 17 days covering Afghanistan and Iran. DFID spent 10 days in India.

[30] Respondents from field programmes were largely those from single-country evaluations.

[31] Jamal, RTE Frequently Asked Questions.

[32] One evaluator commented that information overload was a significant problem, partly because at that stage documentation was unprocessed and not yet prioritised or "cleaned".

[33] While more than half were carried out in rapid-onset emergencies, the Southern Africa RTEs demonstrate its applicability to slow-onset and those in Sierra Leone and Angola to chronic emergencies. UNHCR's last Afghanistan bulletin covers more issues relating to chronic emergencies (Iran) than rapid-onset.

this reason a design template for an RTE should not restrict itself to the relatively rare rapid onset scenario but retain the flexibility to respond to a variety of scenarios.

While experience to date shows a diversity of applications, the implications of wider use of RTE suggest caution. The resources required to implement an RTE mean that few agencies could afford them as a regular tool[34]; neither could they spare the internal staff[35] or find appropriate external evaluators on a regular basis. Established, slower internal mechanisms would normally suffice in regular programmes and an RTE would be less likely to attract interest and therefore support from senior management. These factors suggest that institutionalising RTE as a tool for all programmes would reduce its value and potentially dilute its quality; the temptation would be to use less appropriate staff, compromise on time and timing and ultimately call its utility into question.

2.2.2   The Methodology

Tools
Real-time evaluation methodology and tools differ little from standard evaluations: semi-structured interviews, sample site visits, some in-depth, one-to-one interviews, focus group discussions, stakeholder analysis, and review of secondary documentation.

Participation
The participation of external stakeholders in the RTEs studied tends to be confined to the interview, rather than the feedback phase.

The participation of the field staff is largely similar to a standard evaluation in terms of being available for interviews and support, but with greater interactivity through workshops and discussion fora. The importance of a participatory approach was often emphasised. Facilitating ownership, demonstrating utility and operating successfully in the less-than-ideal environment of the early stages of an emergency appears to be as much a result of the interaction between the evaluators and the staff involved as it is the design of a real-time evaluation. Given that the ability of field staff to absorb lessons and give time to evaluators is conceivably reduced in an emergency context, it is crucial that the evaluator is able to establish an appropriate relationship with the teams and be quickly perceived as constructive and supportive.

Findings and recommendations are, as far as possible, arrived at jointly through a process of discussion and feedback, or achieve a high degree of acceptance prior to departure from the field. Workshops and debriefs are the principle means for more structured participation, but many of the evaluations have relied heavily on informal discussion and a general openness to the views of staff and other stakeholders.

Beneficiary Consultation
Although the shorter time frame of an RTE could preclude adequate consultation with beneficiaries and other agencies, an RTE often spends as much time in the field as do many standard evaluations. In theory, an RTE could emphasise beneficiary consultation, on the basis that it is often weak knowledge of beneficiary views that leads to inappropriate programme design and an RTE could remedy this in good time. In practice, this has not always been the case. Affected by the same conditions as the implementing teams, time, security and logistical constraints are often quoted as understandable reasons for focusing on the staff, rather than the beneficiaries of humanitarian projects. Nonetheless, where the evaluation is of one country, it certainly appears to be feasible to consult more extensively with the beneficiaries. A good example of this is the DEC Goma monitoring mission that, whilst in eastern Democratic Republic of Congo for only a week, employed a local researcher to carry out a series of interviews with beneficiaries and other members of the local population, thus gaining valuable insights (and feedback for the agencies) on local perceptions, analysis and needs.

---

[34] Funding is generally from the evaluation office's budget. Sourcing funding for evaluation early on in a programme may prove difficult, even if a percentage has been included in the programme's budget for evaluation.
[35] UNHCR estimate each RTE takes up a month for a member of staff.

2.3 Managing the Process

2.3.1 Initiating the RTE

Where the evaluation office is designing and promoting the evaluation, it was felt to be essential that collaboration and agreement be established with programme managers. The relationship between the pre-mission preparation and the reception and utility in the field is direct. Interviewees commented that ideally, the TOR should be designed in close collaboration with the country (and regional) offices[36]. Promoting the technique and convincing the regional and country offices of its value to their programme as well as the organisation as a whole appears to be crucial. Given the need to keep an RTE quick and reasonably light on its feet, there must also be a compromise: teams evidently go to the field without the TOR being finalised or completely agreed by all stakeholders. More than half of the RTE TORs studied were drafts and never finalised.

UNHCR emphasises the importance of the evaluators being involved in the response at headquarters, prior to the actual evaluation. In this way the evaluator is fully aware of the themes, communications and thinking that form part and parcel of the response and can also participate in, and influence, the planning process in headquarters. The HAP evaluator was copied in on e-mail traffic before and after the field missions. Some sort of relationship has therefore already been established with the programme and staff.

Establishing the basics of the RTE design prior to an emergency appears to be critical. The process of design, agreeing or obtaining funding, recruiting the team can be lengthy. Two of the NGOs contacted for this study had planned to carry out an RTE, but had run out of time. Either the planner was operational in the emergency and hence already over-committed, or the process of consultation and planning meant that the brief window available for a well-timed RTE had already passed by the time all the elements were in place.

No cases were found of programme, regional or country managers themselves initiating an RTE. This may well be due to the fact that RTE is at an experimental phase in most agencies — not long enough for the approach to become embedded. Additionally, it has tended to be used by agencies that have a centralised structure (e.g. UNHCR and WFP) or those without a permanent managerial relationship (e.g. the DEC, HAP, DFID and IFRC).

2.3.2 Timing

Most RTEs were carried out between six weeks and two months of the start of the response. Fixing a time is less relevant than assessing the phase of the programme; it appears to be a matter of careful judgement according to whether the emergency was rapid onset, chronic, slow, etc. Rapid onset emergencies aimed to deploy an evaluation team "as soon as the dust had settled" — often at around six weeks. Timing in chronic emergencies largely related to the specific programme cycle of the agency (for example key decision-making points in a programme) and the practicalities of recruitment and deployment. Slow-onset, using the Southern Africa examples, could be later if the programme cycle had a slower inception[37].

What is certain is that there is less slack available: the evaluation is changed radically by its timing. Too early, and the evaluators will coincide with the assessment phase and further complicate the logistics of establishing offices, accommodation, etc. Worst of all they could divert and undermine implementing staff at a time when the evaluation's added-value was questionable. Too late and the programme has already made too many irreversible decisions and the RTE can no longer have a real time impact. Groupe URD, faced with the additional burden of fundraising for its Quality Project, deployed to Afghanistan some ten months into the response and faced many of the same problems as standard evaluations would (field staff had already left, etc.).

---

[36] Despite the difficulties in doing so, IFRC went to considerable lengths to ensure that all key stakeholders were in agreement with and had input to the TOR; this included time spent first in headquarters establishing a close link with the relevant operational managers. WFP's first RTE mission, seeking to take advantage of the timing of its Southern Africa response, headed out on a headquarters-generated TOR without time for consultation with the regional and country offices. WFP felt that this reduced the ownership considerably in the field to the detriment of the RTE. Subsequent missions adapted the TORs considerably as a result of field input, including producing reduced and simplified TORs for each mission.

[37] The DEC deploys its monitoring missions normally around six weeks into the response, but delayed this slightly for the slow-onset Southern Africa mission to fit with the progress of implementation.

Many of those interviewed thought that an RTE could be most useful at the end of a first phase emergency period when the programme was entering a rehabilitation or reconstruction phase; the perspective of the evaluators could reduce many of the design errors common to this stage. This has not yet been applied in practice.

2.3.3   The Evaluators

Internal or External
Without exception, agencies sought evaluators who knew its work. They were either internal, though not part of the programme in question (UNHCR, IFRC[38]) or, where they were external, had previous experience of the agency's work, and preferably of the country where the emergency occurred[39]. The ability to grasp internal issues rapidly without recourse to extensive explanation from staff was considered paramount.

If several visits are planned, interviewees suggested that the evaluators should remain the same for consistency and to build on the relationships established.

Size of the Team
Team sizes varied depending on the scale and scope of the evaluation but were mostly between two and three[40]; UNHCR found that one person is highly mobile but could be overloaded and bring insufficient perspective. The only factor that relates to RTE particularly is that the larger the team, the greater the pressure on the country teams to support them, assessed as undesirable during an emergency[41]. RTE planners also anticipated that team size would influence the reaction of those at the receiving end: a large team would be viewed as an imposition rather than a support.

Profile
The profile of the evaluators was considered extremely important. The credibility of the evaluators must be high, being experienced in both field management and evaluation with a slight emphasis on the former. Trust and credibility appear to matter more than usual because of the sensitivity of the situation and because RTE is new and has to earn its place in the organisation's tool box.

In addition, the evaluators need "to be nice". This always helps of course, but several interviewees commented that an insensitive evaluator disliked by the teams could undermine the entire evaluation. Words such as "modest", "understanding", "flexible" "experienced and benevolent" abounded. Particularly during the early stages of an emergency, the evaluator is working with staff who are under considerable pressure and who may not, at least initially, appreciate the presence of an evaluation team. The evaluator must have good interpersonal skills, sensitivity and flexibility, as well as the ability to assert the right to sufficient "evaluation space". One evaluator noted the challenge of exhibiting an understanding of the constraints and stresses, while at the same time maintaining the dispassionate distance required to make recommendations objectively.

Role
The evaluation team may be expected to participate in planning meetings, provide direct advisory input, and even help in establishing systems. The HAP job descriptions explicitly combine an evaluation and advisory role, referring to a real-time evaluator/methodological resource person[42]. The evaluator found that the combination of the two roles was viable.

---

[38] UNHCR uses staff from its evaluation office; IFRC used, for example, a staff member from a Red Cross Society in another region.

[39] The HAP evaluator, while not an ex-HAP employee or consultant, had been extensively involved with the Humanitarian Ombudsman Project and ALNAP.

[40] UNHCR considered a team size of two ideal, WFP, DFID and IFRC mixed internal and external, generalist and specialist to field a team of three, the HAP (a much smaller programme) used one; the DEC varied depending on the size of the response.

[41] UNHCR initially used one evaluator; although this was felt to be less than ideal with respect to workload and perspectives, it had the significant advantage of the evaluator being able to join in with field missions without displacing programme staff or demanding special transport.

As mentioned above, RTE evaluators can play a valuable role in the establishment of monitoring systems and collection of baseline data.

An approach that over-emphasises the evaluator as a resource can reportedly put pressure on him or her to deliver the goods and demonstrate value. Of itself this is no bad thing, but it risks placing the evaluator in a curious position of proving his or her worth, as though he/she were a potential applicant to the team, leaving the evaluator to negotiate the right to evaluate.

World Vision's approach emphasises the role of facilitator rather than evaluator; similarly Groupe URD demand "proven teaching and communication skills". While neither defines their approach as an evaluation, it is noticeable that agencies using RTE, such as WFP, are similarly increasingly emphasising facilitation skills.

Recruitment
The agency's head office, usually the evaluation office in agencies large enough to have one, had recruited the evaluator(s) in all cases studied. IFRC's evaluation office proposed the team in collaboration with the Operations (Programme) Manager; WFP's evaluation department recruited its team, but also responded to requests for adapted team composition from the field. UNHCR use staff from its evaluation office.

The timely recruitment and deployment of the evaluators is essential to the utility of an RTE. Too late and too much and the value could be lost. The use of external consultants may increase the risk of late deployments through lengthier procedures and lack of availability of the consultants[43].

2.3.4   Follow-up

Reports
The length of reports and the speed with which they are disseminated is an important distinguishing feature of RTE from standard evaluations. The approach is fairly consistent: a brief report (two to five pages) is produced while still on mission, accompanied by a debrief or workshop in the field and subsequently in the regional and head offices. A longer report (10 to 40 pages) is produced for the purposes of broader institutional learning and accountability between four days and two weeks of the field mission.

If the CO is the primary end user of an RTE, and no official follow-up mechanisms exist, the reports are relatively low in the learning hierarchy. RTE effects learning through interactive, not written fora. It is the workshops and discussions that effect change in the field; the brief mission reports disseminate information upward in the organisation, to senior managers and decision-makers. The longer reports contribute to institutional learning.

The common process of producing a series of reports from field to final also lends itself to making appropriate and rapid adaptations vis-à-vis different audiences and accountability needs.

The important points made by interviewees are that the reports should be relatively short and rapidly available, during the programme and in time to make a difference. These characteristics can be adapted according to the size of the programme and the speed of change. It was commented that there is no reason to stick rigidly to a maximum length and turnaround for all emergencies; few emergencies are so fast-moving that the difference of a week would mean that the utility of an RTE was lost.

Follow-up Mechanisms
In general, the mechanisms for implementing RTE recommendations are either the same as for standard evaluations or are opportunist. In headquarters, an RTE may produce a specific and additional mechanism for immediate action, such as UNHCR's Deputy High Commissioner calling an impromptu meeting of regional representatives in response to the Sudan-Eritrea report, or IFRC holding a presentation of findings in Geneva to key stakeholders. These mechanisms are a response to each particular RTE and are as much a function of the size and profile of the emergency as of the RTE itself. The chief characteristic of follow-up is

---

[43] The timing for example of the HAP RTE to Afghanistan may have weakened its value — the first mission was curtailed and the second delayed and coincided with the Trial closing down.

its speed of dissemination and access to decision-makers. Arguably, in high-profile emergencies, an RTE can contribute to an accelerated learning process within the organisation by capturing lessons at a time when all eyes are turned to that programme.

None of the agencies interviewed had introduced a formal follow-up procedure to an RTE evaluation even if, as in the case of WFP, standard evaluations are routinely linked to a management response matrix[44]. This appears to be partly owing to sensitivities regarding the introduction of "yet another" mechanism to the country offices.

According to interviewees, many of the RTE recommendations were acted upon. The second WFP mission noted that many of the action points of the first mission had been implemented.[45] IFRC commented that the RTE was "useful beyond expectation"[46] with significant corrective action taken already in its Southern Africa programme. UNHCR note several changes as a result of its RTEs, including for example improvements in its procurement procedures as a result of findings during the Pakistan RTE[47].

Implementing Recommendations in the Country Office
Given the lack of formal systems it is perhaps surprising that agencies report such success in the implementation of changes. This is partly because most of the impact reported relates to changes within the country programme itself rather than organisation-wide reform. Change is then related to lighter decision-making structures and a higher degree of participation. All the RTEs examined held in-country debriefs or workshops prior to departure. Additionally, the more participatory nature of the RTEs studied mean that discussion and interaction had taken place in the field before the report was finalised. Agreement or otherwise had already have been clarified between the evaluation and the country representative and team. Recommendations were felt more likely to be "owned" by the implementing teams and managers because they a) reflected the view of the field staff, b) were taken as constructive advice or c) were the product of discussions and reflection leading to shared views. Recommendations are also, theoretically at least, more likely to be realistic and do-able as a result. Country offices can normally rapidly implement changes within their control without lengthy procedures and through the usual in-country management systems. Indeed evaluators commented that many changes were implemented during the mission, prior to the feedback or report.

Senior Management Support for Change
Whilst there are no apparent formal mechanisms in place for RTE follow-up in agencies studied, powerful informal factors that relate to senior management ownership may explain the apparent success. Regardless of its coyness about compliance-driven mechanisms, a UNHCR evaluation is explicitly supported by senior management. The team, like EPAU, reports to the Assistant High Commissioner and everybody knows it. By contrast, the mechanisms for follow-up to the DEC monitoring missions are more diffuse and decentralised due to the nature of the DEC and the number of different agencies involved. Final DEC evaluations have commented on the lack of implementation of monitoring report recommendations. This may be due in part to the absence of any mechanism, formal or informal, to facilitate follow-up. It may also relate to the lack of a "champion" who sees it through distribution, meetings, "after actions" and other "follow-up".

No special mechanisms were found related to organisation-wide learning. RTE lessons feed into the organisation's knowledge management in the same way as standard evaluations: recommendations go to the Board, to Steering Committees, to Oversight Committees and so on. The chief difference appears to be the result of the profile of the emergency: a real-time evaluation is more likely (though not necessarily) to be

---

[44] UNHCR's approach to evaluation in general is not compliance-driven. The IFRC and DEC have a complex accountability mechanism within a federal structure that does not mandate the secretariats to enforce recommendations in an audit-like fashion. Proposals for a permanent HAP mechanism will establish the extent to which RTE recommendations have been acted upon. More generally, this study can only report on the mechanisms that exist to support an RTE, not on their follow-up and impact.
[45] Communication with Jon Bennett, consultant real-time evaluator for WFP Southern Africa RTE.
[46] Communication with Mathew Varghese, IFRC.
[47] It is difficult to follow findings and recommendations from one RTE mission to another in the Afghanistan context as the locus of the UNHCR response and of the RTE missions themselves changes on each occasion. This is an inherent risk in evaluating the impact of RTEs.

conducted during a large, high-profile emergency and hence feed into organisational processes established for significant evaluations[48].

Management Lines and Decision Making

The proximity of the real-time evaluators to programmes and the immediacy of the recommendations mean that evaluation and the programme activities are inextricably linked; thus how decisions are taken as a result could lead to confusion in the field. Some difficulties were reported when the evaluator encountered management weaknesses and dysfunctional teams and when the evaluator was expected to deliver solutions on the spot. In both cases this was partly due to a lack of clarity in the TOR regarding the role of the evaluator — as one who identifies the problem or as a trouble-shooter who fixes it. Clarifying the role of the RTE team, how recommendations would be implemented, lines of authority related to decision making and the levels of authority required depending on the type of recommendation or action point would be important prior to deployment; resolving the dilemma should not be the job of the evaluator(s). That said, in most cases interfering with or confusing management lines does not appear to have been a problem. The *process* of observation-advice-discussion in the field largely led to field managers simply absorbing action points in their everyday management processes.

The further from the field, the harder it is to identify decision-making mechanisms related to RTE. If rapid-response management mechanisms do not already exist, then it appears that an organisation will either have to invent them, initiate an extraordinary mechanism specific to the response or adopt an informal approach that may take advantage of existing mechanisms, such as EPAU's participation in its head office crisis cells.

3.   A SUMMARY OF THE STRENGTHS AND WEAKNESSES OF RTE

RTE has the same principle goals as a standard evaluation and its value-added largely relates to its timing, as do its weaknesses.

*Advantages***:**

All agencies interviewed planned to continue to use RTE for the following reasons:

To the Programme Being Evaluated:
- Its recommendations can be used "at a time when key operational and policy decisions are being taken…in time to make a difference"[49];
- The evaluators bring perspective and objectivity at a time when, typically, the pressure for service delivery displaces strategy and reflection;
- Its enhanced understanding of the context can lead to more appropriate and realistic recommendations;
- It facilitates in situ "learning by doing, not by injunction"[50], which may be a highly appropriate form of learning for field staff;
- The importance given to interaction and adequate debrief means that recommendations pertaining to the field are more likely to be realistic and are essentially already agreed, circumventing disagreement and lengthy "sign off";
- It supports greater accountability through learning and has the potential for greater accountability to external stakeholders if reports are published;
- It can facilitate team-building, boost morale, surface, and at times resolve, tensions;
- It links the views of different (internal) stakeholders, particularly bridging the gap between the country office, regional office and headquarters, bringing up-to-date information from one to another and acting as "a voice" or advocate when appropriate.

---

[48] For example, UNHCR now establishes a Steering Committee for all major evaluations, which presents its findings to a Senior Management Committee. This mechanism is too slow for the immediate-impact findings of an RTE, but can serve considerations for policy development in the longer term.
[49] Jamal, RTE Frequently Asked Questions.
[50] Broughton.

To Organisation-wide Learning:
- It enhances a standard evaluation through capturing unique information often unavailable to standard evaluations such as issues related to process, relationships and the experience of testing policy in practice;
- It mitigates the impact of lost learning through high staff turnover;
- It can bridge the gap between monitoring, review and evaluation, bringing additional objectivity and weight to its recommendations;
- It can facilitate the establishment of appropriate baseline surveys and monitoring systems, improving impact assessment during and upon completion of the programme;
- It can use the opportunity afforded by the increased attention within an organisation focused on a high-profile emergency to accelerate learning.

Risks and Deficiencies:

Many of the shortcomings of RTE are associated with it being done badly, rather than with it being done at all. Several shortcomings could equally describe standard evaluations. However, given the timing of an RTE, often in an intense early stage of an emergency, the disadvantages have a potentially greater negative impact. It should be noted however that many of the following are perceived risks, not those that necessarily occurred.

Overall:
- The opportunity costs of field staff time exceed the value of the RTE;
- Damaging the agency's reputation and relationships and undermining its ability to function;
- A snapshot of a particular moment in time that may not be representative of the programme and context.

Design:
- The risk of superficiality and/or an unrealistic scope of the TOR in the limited time available;
- Weaker analysis due to a narrower focus on the agencies' own programme and limited (analytical) secondary documentation.

Timing:
- Staff's ability to work with the evaluator(s) and absorb lessons is limited due to time pressure and stress;
- The findings could generate a perception (positive or negative) of performance that sticks regardless of changes and subsequently affects funding, morale, motivation or complacency;
- The margin of error is very narrow: delays could neutralise the real-time advantage completely.

Human Resources:
- A confusion regarding the role of real-time evaluators who will often be called upon to participate in programme planning and even expected to trouble-shoot and provide solutions to the detriment of their 'evaluator' role;
- Timely availability of (appropriate) staff to carry out the RTE is likely to be limited, yet the choice of staff is critical.

Process:
- The advantage of the evaluator's proximity to process and context turns to disadvantage through a loss of perspective. Evaluators are susceptible to overemphasising ephemera or being influenced by inaccurate collective beliefs held at the time of the evaluation;
- Greater reliance on interviews increases the risk that evaluators are overly influenced by gossip or the strongly held views of a few individuals;
- Badly managed, the process could damage rather than improve morale;
- The need to implement rapidly can prevent adequate consultation between key stakeholders, reducing ownership and the inherent value of the RTE's interactive approach.

Management Mechanisms and Decision-Making:
- Confusing decision-making and management lines through the implementation of RTE action points *in situ.*

Good Practice:

The following is a summary of suggested good practice drawn from the experience in real-time evaluations so far. In some cases the "good practice recommendations" are not existing practice, merely that proposed by agencies on the basis of experience so far.

Design:
- A full process of consultation with key (internal) stakeholders in the TOR;
- Clarity of purpose prior to implementation dictating:
  - the roles of the evaluators (facilitation, resource, advisor, joint planner),
  - the balance between learning and accountability (and hence report length and dissemination mechanisms),
  - to whom the evaluation is accountable, internal and external, upwards and downwards,
  - follow-up mechanisms (who and how at each level);
- Ensuring adequate consultation of beneficiaries;
- Creation of generic TORs previously discussed with all regions or country offices, enabling the rapid production of a short, accessible TOR for a particular phase of the programme;
- Responsiveness to the phase of the programme and emerging issues.

Management:
- Advance preparation by the RTE initiators with respect to recruitment, funding and design, maximising the likelihood of timely implementation and team deployment;
- Securing the support of senior management at inception;
- Multiple methods of reporting: immediate verbal reporting combined with short preliminary reports or bulletins, followed by longer reports for wider circulation and longer-term issues.

Evaluators:
- Small evaluation teams composed of experienced personnel very familiar with the organisation's work;
- Evaluators who are involved with the programme prior to the mission (through participation in key meetings or through being part of key communications);
- An evaluation team that remains consistent throughout a series of missions on the same RTE.

Participation and Interaction:
- Participatory approach in the field, facilitating reflection and learning in the team and capturing findings and action points through an interactive process;
- Adequate debriefing in the field of an interactive nature rather than a presentation, enabling ownership and rapid implementation of recommendations;
- Using the evaluation to review and, if appropriate, improve monitoring systems and baseline data collection.

4.   NEXT STEPS FOR UNICEF

It is suggested UNICEF consider the following steps:

- Carry out a simple, brief stakeholder analysis of the potential "users" of an RTE;
- Plan a pilot in complete collaboration with key stakeholders, particularly a Regional Office (RO) to establish the overall purpose and design;
- Create a generic TOR;
- Decide prior to implementation whether the RTE reports would be published and design the versions of reports appropriately, giving brief guidelines to the evaluators;
- Obtain high-level support for RTE before initiating even a pilot. This would greatly facilitate support for tackling weaknesses and institutional problems beyond the capacity of the deep field to influence;
- Agree on decision-making authority prior to implementation and build into the TOR;
- Agree on the mechanisms for collaboration between the evaluation office and programme management, including exactly what role the evaluation office would play in an RTE;
- Draw up a roster of real-time evaluators in collaboration with the regional offices;
- Obtain funding, or at least the agreement of funding, for a pilot RTE;
- Agree with the highest level of authority possible, a clear and transparent mechanism of upwards accountability. This would greatly enhance reception and perceived value of an RTE in the country offices;
- Consider benchmarks that will facilitate the inclusion of UNICEF's partners, such as the use of Sphere standards likely to be used by partner NGOs;
- Build the results of a final RTE into existing mechanisms such as the mid-term review.

ANNEXE I

FURTHER DISCUSSION ON DEFINITION AND DESIGN

1. COMPARING RTE TO OTHER PROGRAMME MECHANISMS

Informal definitions tend to classify a process involving solely internal staff, carried out during an operation and unpublished as a review and a mechanism involving at least one external member carried out towards the end or following completion as an evaluation, published or not. UNHCR to date has exclusively used its own staff and indeed refers to RTE as an "interactive peer review"[51]. Additionally, a key characteristic of RTE, "that it is conducted during implementation rather than at completion or ex post and not necessarily as a one-off event"[52] lends itself more to the concept of review than of evaluation. However, the literature on evaluation does not dismiss the use of internal teams[53] and, although characterised by shorter reports than most standard evaluations, all RTEs[54] identified for this review fulfil ALNAP's criteria and characteristics of (standard) evaluation, being systematic, impartial, intended to support lesson-learning and accountability, utilising recognisable evaluation criteria and providing outputs in the form of written findings and recommendations[55].

Still, by this description many organisational reviews would be classified as evaluations. A pragmatic way of differentiating could be simply that a team directly involved with implementing a programme cannot (be assumed to) be sufficiently dispassionate or disinterested to achieve the impartiality demanded for an evaluation. Therefore, regardless of how systematic the approach, a mechanism implemented by the implementers would be termed a review.

None of the RTEs assessed that employed internal teams utilised staff responsible for those programmes; internal staff were drawn from other programmes and from the evaluation unit of the organisation; internal, but without being in the firing line.

RTE similarly shares part of the terrain with monitoring. The current study commissioned by ALNAP on monitoring refers to monitoring as "the systematic and continuous assessment of the progress of a piece of work over time…a basic and universal management tool for identifying the strengths and weaknesses in a programme. Its purpose is to help all the people involved make appropriate and timely decisions that will improve the quality of the work"[56]. Goyder's definition of Strategic Monitoring to "establish whether the overall strategy on which a programme is based is appropriate" lifts monitoring out of its common control and output focus and encompasses broader considerations for change in a dynamic setting — akin to RTE and bordering on evaluation in general. Nonetheless RTE is not monitoring. It is intermittent, not continuous and, unlike monitoring, is not carried out by internal staff for internal information purposes.

2. WHAT PURPOSE RTE?

Extracts from TORs:
UNHCR: "its broad objective is to gauge the effectiveness and impact of a given UNHCR response, and to ensure its findings are used as an immediate catalyst for organisational and operational change. […] UNHCR also hopes to reinforce the link between operations, evaluation and policy formulation" [57] and: "The findings of the RTE are intended to serve as a tool to assess and modify, as appropriate, UNHCR's response to the current crisis, and to assist in longer term learning and emergency preparedness capacity"[58].

---

[51] Jamal. Real-time humanitarian evaluations. Some frequently asked questions. UNHCR 2002
[52] Broughton.
[53] For example, ALNAP's defining characteristics of evaluation include the use of internal teams. ALNAP Annual Review 2002 p201
[54] For this reason, of the reports studied, UNHCR's real-time evaluations are the closest to what could be referred to as a review, as the analysis uses UNHCR's policies as its benchmarks without necessarily referring to standard evaluation criteria.
[55] ALNAP defines evaluation as "A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability" and offers a set of characteristics covering commissioning, team composition, evaluation criteria and reporting. ALNAP Annual Review 2002 p201
[56] From Gosling & Edwards (1995) quoted in Hugh Goyder Report No 1. Scoping Report in Strengthening Programme Monitoring Systems in the Humanitarian Sector.
[57] Jamal, RTE Frequently Asked Questions.
[58] *TOR Real-time evaluation of UNHCR's response to the Eritrea-Sudan displacement emergency* (May-June 2000).

WFP:
"This real-time evaluation of WFP's response to the Southern Africa crisis will be the first exercise of its kind within WFP and is expected to yield valuable insights into the workings of the operation as well as substantial and lasting learning benefits for the organization as a whole[59].

HAP:
Sierra Leone:
"The purpose of the real time evaluation team is to provide an external perspective on the validity of the HAP's self-assessments. It will fulfil this purpose by playing a double role:
• Acting continuously as a 'critical sounding board'…thereby enhancing the internal reflection and learning among the HAP staff, and
• Provide constructive and practical guidance and support during the field research, to help make it as effective as possible[60].

and in Afghanistan:
"The two main purposes of the RTE are:
(i) …To provide constructive and practical guidance and support during the field research, to help make it as effective as possible, to provide recommendations to HAP Geneva regarding potential course corrections,
(ii) …To provide an external and evidence-based observation of the work done at field level by the HAP … identifying recommendations and findings for the remaining HAP field trials and for a permanent field mechanism of accountability [61].

IFRC:
Purpose: "To provide management and decision makers with a timely situation analysis of progress to date, and to identify and facilitate any necessary modifications that may be required to inform this on-going operation".
1. Provide managers and decision makers with information on the relevance and effectiveness in achieving the objectives of the emergency response;
2. Provide managers and decision makers with a situation analysis of the challenges (contextual, operational etc) that exist;
3. Suggest relevant responses and necessary corrective action to these challenges for consideration by decision makers[62].

DEC:
• "Review impact of the early phase of the response – the reduction of mortality, morbidity and suffering in the early phase
• Advise on programme adjustments
• Review the strategic use of the DEC fund and the division of labour and resources between the Members"
It has been further highlighted in previous evaluations that given staff turn over and a rapidly evolving situation an early mission is important to feed information into the main evaluation".[63]

3. DESIGNING REPORTS FOR RTE

Broughton recommended that WFP produce an Aide Memoire of around 5 pages whilst still on mission, a summary report of a maximum of 5,000 words (@10 pages) and a full report of up to 20,000 words, including annexes. The actual Southern Africa TORs similarly demand an Aide Memoire, followed by a (longer) 20-page interim report ("@10,000 words). The final report following the three field missions will be a shorter, 10-page report.

---

[59] Draft Terms of Reference for a "Real-Time" Evaluation of WFP's Response to the Southern Africa Crisis, 8 July 2002 thru 31 July 2003.
[60] HAP. Job Description of Real-time Evaluation Sierra Leone, Humanitarian Accountability Project 2001-2002.
[61] *Job Description Real-time Evaluator/Methodological Resource Person*. Humanitarian Accountability Project Afghanistan 2002.
[62] *Real-Time Evaluation of the response by the International Red Cross and Red Crescent Societies to the Southern African emergency*. Terms of Reference. Draft. IFRC 2002
[63] Draft Terms of Reference For Southern Africa Crisis Monitoring Mission. Disasters Emergency Committee (DEC), 2002

IFRC's team were asked to produce a report for each country visited prior to departure of no more than four to five pages, followed by a consolidated report no longer than ten pages. The actual reports were around two to three pages per country and the full report around 12, excluding the executive summary.

DEC monitoring reports are longer, the two viewed being between 16 and 25 pages, excluding executive summary. The draft is expected a week following the mission.

DFID's RTE to India produced a "debriefing note" within four days of the field trip and a report of around 20 pages.

HAP produced short, four- to five-page reports following the first field missions and final reports of 30 to 40 pages. The latter were produced some time after the completion of the Trial owing to sensitivities regarding content; immediate learning was intended to occur through interaction in the field. The reports lend themselves to the more traditional evaluation utility of learning for the future.

UNHCR produces short reports ("bulletins") of 10 to 15 pages. The first report of the Sudan-Eritrea pilot was around 20 pages, plus an executive summary; subsequent reports are shorter. A first draft version is produced in the field and subsequently used as the basis of debriefs or presentations in Geneva; the bulletin follows within a week or two for wider distribution.

ANNEXE II

References

Bhatt, M and M. Reddick. *Developing the Learning Office Concept*, ALNAP, May 2000.

Broughton, B., *Proposal Outlining a Conceptual Framework and Terms of Reference for a Pilot Real-Time Evaluation,* OEDE/WFP, 24 December 2001.

Christoplos, I., *The Humanitarian Accountability Project, Field Trial in Afghanistan, May 2002-July 2002*, Humanitarian Accountability Project, October 2002.

Christoplos. I., *Evaluation Report. The Humanitarian Accountability Project, First Trial in Sierra Leone, November 2001 to May 2002,* Humanitarian Accountability Project, 16 July 2002.

Crisp, J. and E. Stigter, *Real-time Evaluation of UNHCR's Response to the Afghanistan Emergency. Bulletin No. 2,* EPAU, UNHCR, 2002.

Disasters Emergency Committee, *Draft Terms of Reference for Goma Volcano Monitoring Mission*, DEC, 2001.

Disasters Emergency Committee, *The Independent India Earthquake Evaluation: Draft Terms Of Reference India Earthquake Monitoring Visit 11-25 March 2001,* DEC, 2001.

Disasters Emergency Committee, *Draft Terms of Reference for Southern Africa Crisis Monitoring Mission,* DEC, 2002.

DFID, *Terms of Reference Real Time Assessment of Federation Response To India Floods*, 2001.

ETC UK, *Real Time Evaluation of the Humanitarian Response to the Crisis in Kosovo,* Danida, May 1999.

Goyder, H., *Strengthening Programme Monitoring Systems in the Humanitarian Sector: Report No. 1: Scoping Report*, ALNAP, 2002.

Groupe URD, *Proposal for an Innovative Evaluation Procedure. Iterative Evaluation Process with Mini-Seminar Combination*, Groupe URD, October 2000.

Groupe URD, *Le Projet Qualite en Afghanistan: Mission 1: 20 Juillet 2002 - 24 Aout 2002,* Groupe URD, November 2002.

Groupe URD, *The Quality Project In Afghanistan : Mission II. Terms of Reference First Draft*, GROUPE URD, 22 December 2002.

HAP, *Job Description of Real-time Evaluation Sierra Leone, Humanitarian Accountability Project 2001-2002, HAP, 2001.*

HAP, *Job Description Real-time Evaluator/Methodological Resource Person,* Humanitarian Accountability Project, Afghanistan 2002.

IFRC, *Real-Time Evaluation of the response by the International Red Cross and Red Crescent Societies to the Southern African emergency*, IFRC, 2001.

IFRC, *Southern Africa Food Security Operation — Real Time Evaluation Report*, November 2002. IFRC, 2002.

Jamal, A., *Angola 2000. A Real Time Assessment of UNHCR's IDP Intervention,* EPAU, UNHCR, 2000.

Jamal, A., *The Sudan/Eritrea Emergency May-July 2001, An Evaluation of UNHCR's Response*, EPAU, UNHCR, February 2001.

Jamal, A., *Real-time evaluation of UNHCR's response to the Afghanistan emergency. Bulletin No. 3.* EPAU, UNHCR, May 2002.

Jamal, A. and J. Crisp, *Real-time Evaluations, Some Frequently Asked Questions,* UNHCR, EPAU, May 2002.

Lockwood, H., *Real-Time Assessment of The IFRC Response to The India Floods*, DFID, October 2000.

Macaulay, S., *The Volcano Eruption in Goma, Democratic Republic of Congo Report of a Monitoring Visit,* Disasters Emergency Committee, April 2002.

Raynard, P., *Mapping Accountability in Humanitarian Assistance,* May 2000.

Reddick, M., and J.Telford,. *Further Development of the Learning Office Concept With Reference to Sierra Leone and East Timor and a Developed Proposal to Run a Test Office during 2001. For Discussion at the 8th ALNAP Biannual Meeting*, Geneva, 19-20 October, 2001.

Stigter, E. and Crisp, J., *Real-time evaluation of UNHCR's response to the Afghanistan emergency, Bulletin No.1*,EPAU, UNHCR, November 2001.

UNHCR, *Real-time evaluation of UNHCR's response to the Eritrea-Sudan displacement emergency (May-June 2000)*, EPAU, UNHCR, 2000.

UNHCR, *Terms of Reference: Real-time evaluation mission to Pakistan 9 - 24 November 2001*, EPAU, UNHCR, 2001.

UNHCR, *Terms of reference, Real-time evaluation, part 3: Iran and western Afghanistan. Draft terms of Reference*, EPAU, UNHCR, 2002.

van de Putte, B., *Follow-up to Evaluations of Humanitarian Programmes, Paper submitted to the ALNAP Biannual meeting, 26-27 April 2001.*

Vaux, T., *The Earthquake in Gujarat, India. Report of a monitoring visit for the DEC,* DEC, March 2001.

WFP, *Draft TOR for "Real-Time" Evaluation of WFP's Response to the Southern Africa Crisis Second mission: 17 November-22 December 2002.* OEDE, WFP, October 2002.

WFP, *Draft Terms of Reference for a "Real-Time" Evaluation of WFP's Response to the Southern Africa Crisis 8 July 2002 thru 31 July 2002.* OEDE, WFP, 2002.

ANNEXE III

People Interviewed

John Mitchell	ALNAP Secretariat
Ian Christoplos	Independent Consultant
Hugh Goyder	Independent Consultant
Francois Grunewald	Groupe URD
Jeff Crisp	Head of EPAU, UNHCR
Scott Green	Evaluation Officer, WFP
Julian Lefevre	Chief Evaluation Officer, WFP
Simon Lawry-White	Independent Consultant
Brendan Gormley	Chief Executive, DEC
Moira Reddick	Independent Consultant
Tony Vaux	Independent Consultant
John Telford	Independent Consultant
Mathew Varghese	Head of Planning, Monitoring and Evaluation, IFRC
Rupen Das World Vision	Humanitarian Assistance Team Leader, World Vision
Maurice Herson	Monitoring and Evaluation Advisor, Oxfam GB
Alistair Hallam	Valid International
Bernard Broughton	Independent Consultant
Adrian Denyer	Emergencies Advisor, CARE UK/International
Jon Bennett	Independent Consultant
Arafat Jamal	Evaluation Officer, UNHCR
Guy Ouellet	Representative, UNHCR Angola
Surendra Panday	(previously) Head of Sub-Office, Iran
Elizabeth Eyster	Executive Assistant to the Representative, Islamabad, UNHCR
Andy Featherstone	(previously) HAP Team Leader, Afghanistan
Simon Springett	Humanitarian Co-ordinator, Oxfam GB Zambia

ANNEXE IV
1 November 2002

Terms of Reference
Desk review of real-time evaluation experience

Background
In emergencies more than ever, managers are challenged to keep in focus the overall purpose and objectives of programming and advocacy work, the status of implementation in relation to objectives and the changes in the external environment that determine the relevance of both the overall purpose and the implementation strategies adopted. Monitoring and evaluation systems are the key to ensuring that programmes are relevant and that the expected results are being achieved. Yet, in a process of reviewing UNICEF humanitarian response capacity, monitoring and evaluation at CO level has been identified as a weakness. In crisis and unstable contexts, pressures for immediate programme implementation draw staff and management attention away from monitoring and evaluation functions, both in the short-term immediate humanitarian response and as an emergency continues. Where COs do have strong M&E staff they are typically absorbed in monitoring the situation of women and children, especially where traditional information systems collapse. Staff who are over-stretched and pressured have difficulty in stepping back for critical analysis. Further, the higher numbers of stakeholders and heightened emphasis on co-ordination, including in-house between programme and operations staff, renders evaluative exercises more complex and potentially more volatile and/or political.

In response to this analysis, the UNICEF Evaluation Office (EO) and Office of Internal Audit (OIA) are pursuing two strategies to strengthen CO M&E capacities. One is to boost the M&E skills and knowledge of staff in COs in crisis and unstable contexts through targeted on-the-job training. The other, the focus of this TOR, is to establish a "real-time" evaluation mechanism. This is all being undertaken in the context of an organisation-wide process to strengthen UNICEF humanitarian response capacity supported by a programme of co-operation with the UK's Department for International Development.

The proposed "real-time" evaluation mechanism is understood to mean:
- A mechanism (management structure, procedures, evaluation methodology and tools, resources) to provide direct support to COs initially managed from headquarters but ultimately more likely from Regional Offices, at the early stages of a humanitarian response, ideally within the first 3 months[64].
- Evaluations that span programme design, operational strategies and management systems as well as practice supporting programmes, including the establishment of monitoring and evaluation systems.
- An evaluation approach that provides assessment of CO performance and actually provides or facilitates immediate hands-on support to COs to address problems identified.
- Evaluations that are quick and light, but systematic and well structured.

There is experience outside of UNICEF in using real-time evaluation. UNHCR and WFP have both advanced thinking and experience in this area. Non-governmental organisations (NGOs) have also advanced work in this area (e.g. Groupe URD). Prior to developing a real-time methodology, tools and overall mechanism, it is essential to review the existing experience to establish some of the important practices to be adopted and/or questions to be resolved.

Purpose
The desk review is the first step in a multi-year process of developing and testing a real-time evaluation methodology and institutionalising this as part of UNICEF's evaluation system. The desk review will provide the basis for the next stage of work, the development of a draft methodology and field-testing.

The desk review is therefore intended to identify key lessons on real-time evaluation of humanitarian action drawing from a comparative analysis of the experience of other agencies and organisations in this area. Lessons should include positive practices to emulate or challenges to be resolved regarding the real-time evaluation mechanism, methodology and tools.

---

[64] As indicated below, the ideal timing of the real-time evaluation needs to be considered carefully in light of experience and is likely to differ depending on the type of emergency and the nature of UNICEF response.

The desk review will be used by UNICEF EO/OIA to generate and structure an initial discussion of options in the design of real-time evaluation mechanisms and methodology involving key stakeholders within the organisation.

Scope

The following is an initial outline of issues to be explored in the Desk Review. This listing will be revisited after the consultant has undertaken a preliminary quick scanning of documentation and a first round of interviews at UNICEF headquarters.

Overall use

- How have the findings, recommendations and lessons of real-time evaluations been used at different levels of an organisation and what aspects of the mechanism, process design or methods have contributed to good reception of evaluation results?
- What is the added value of real-time evaluation in its different forms?

Managing a real-time evaluation mechanism

- How should the real-time evaluation be *initiated*? By whom? In what conditions? How long after the beginning of a crisis should it take place (and how is this defined)? To what degree is this influenced by the type of presence that the organisation has (i.e. long-term vs. emergency response only)?
- Who *recruits* the evaluator/evaluation team?
- What *management structure* is best suited?
- How are real-time evaluations *financed* and what is the *cost*?

Design of the real-time evaluation methodology

- How is the *objective* of the real-time evaluation best defined?
- Who is the *primary end-user*?
- What should be the *scope* of the evaluation?
- To what degree can/should the real-time evaluation address *standard evaluation criteria* for humanitarian assistance: relevance, efficiency, effectiveness, impact and coverage, sustainability/connectedness, co-ordination, coherence, protection? How is this best done?
- To what degree and how can the real-time evaluation address *assessment of both programme design and management systems/practice, including monitoring and evaluation planning, systems and their actual use*?
- To what degree and how can the real-time evaluation in a given CO address both *assessment and the provision or facilitation of immediate support* to address weaknesses identified?
- What *methods of data collection and analysis* are best used?
- What is the ideal *participation* of the CO staff — who, how much time, what role?
- What is the ideal *participation* of external stakeholders (counterparts, partner organisations, donors) -- who, how much time, what role?
- What is the ideal *evaluator/evaluation team profile/size*?
- What *limitations* to the evaluation must be accepted?
- How does all of the above vary according to the *timing* of real-time evaluations?

Real-time evaluation tools

- What issues can be covered in the evaluation with standardised assessment tools?
- To what degree can tools also serve as self-assessment checklists to guide CO managers?
- How well do/can tools accommodate the evolving nature of emergency programme response?
- What existing tools, elements of existing checklists and/or evaluation questions are most promising for UNICEF to explore?

In addition to tackling the above questions, the desk review should also involve identifying and obtaining any established tools, and presenting an analysis of those considered promising or with promising content. This analysis can be presented in an Annex.

Existing information sources
The documentation already identified for the desk review includes concept pieces describing real-time evaluation, terms of reference for actual real-time evaluations and real-time evaluation reports. Several UNHCR documents on its real-time evaluation experience are available on the Internet. An initial small selection of grey literature has been made available to UNICEF by both UNHCR and WFP. A more in-depth search for documentation will be one of the first tasks in undertaking the desk review.

The analysis in the desk review will require reference to UNICEF evaluation policy, methods and tools, in particular experiences in Country Programme Evaluation. Relevant reference materials will be available on the UNICEF Internet website as of 16 September (at www.unicef.org/reseval). Similarly, the analysis will require reference to draft materials on UNICEF OIA Audit Guidelines for the Assessment of Essential Controls in Emergency Programmes.

Methodology
While called a "desk review", this exercise will necessarily combine documentary review with telephone interviews of key informants in those agencies/organisations most advanced in the experience of real-time evaluation (in particular UNHCR and WFP).

The tasks to be undertaken include the following:

- Initial review of documentation already identified. Establish an awareness of immediately available literature on real-time evaluation as well as of UNICEF evaluation policy, methods and tools.
- Briefing visit to UNICEF NY (2 days). Revisit and clarify the Terms of Reference as necessary, in particular defining the scope of the desk review. Clarify any issues coming out the initial review of documentation. Carry out initial interviews with key UNICEF staff to help orient the documentary review. Prepare an outline of the structure of the final Desk Review report and other products.
- Finalise key questions for analysis. Either during or immediately following the Briefing visit, revise and finalise the key questions for analysis as initially outlined in the Terms of Reference.
- Carry out further documentary search and identify key informants Systematically identify UN agencies, international non-governmental organisations and any other sources with experience in real-time evaluation, identify and obtain relevant documentation and identify individuals with direct experience in management or implementation of real-time evaluation as potential key informants. UNICEF will facilitate contacts with key organisations, as well as use of the Active Learning Network on Accountability and Performance in Humanitarian Action "listserve" to generate contacts.
- Carry out key informant interviews Finalise selection of key informants. Carry out interviews via telephone.
- Draft desk review.
- Prepare and submit first draft, including appendices with analysis of existing tools.
- Discussion/review of first draft.
- Clarify comments of UNICEF reference group on first draft.
- Finalise desk review.

The desk review will be limited by the availability of documentation. The number of agencies/organisations with significant experience with real-time evaluation mechanisms is expected to be small.

The calendar for the above activities is flexible. The following is an indicative calendar of key milestones.

Milestones	Date
Briefing visit to NY	1st half November*
Discussion/revision of key questions	2nd wk November*
Submission of initial draft for comment	20 December
Submission of final draft	16 January
*dates flexible	

Qualifications
The consultant must be at a senior level, with evaluation experience in the humanitarian assistance sector, in particular at the level of the overall humanitarian response of one agency/organisation. Experience with meta-evaluations is also critical. S/he must also have relevant knowledge and experience of programme management in the context of humanitarian assistance. The consultant must have excellent synthesis, analytical and written communication skills. Familiarity with UNICEF is a strong advantage.

Accountabilities
The consultant is responsible for carrying out each of the above-mentioned tasks and producing the deliverables as described below. The consultant will report to the Evaluation Office.

A Reference Group, formed by the Evaluation Office, including representatives from the divisions of Emergency Programmes, Planning and Policy, and Programme and the Office of Internal Audit as well as selected individuals from regional offices will comment on the first draft of the desk review.

The Evaluation Office will be responsible for:
- Providing relevant documentation on UNICEF evaluation policy;
- Planning and co-ordinating the Briefing visit;
- Approval of finalised Terms of Reference including any revisions to the scope of the desk review as well as the final structure of the products;
- Facilitating access to networks and introduction to individual partner agencies as necessary;
- Co-ordinating and consolidating the comments of a Reference Group on the first draft of the final report;
- Approval of the final products submitted.

Procedures and logistics

Outputs
On completion of the contract, the consultant will submit the following to Evaluation Office:
- Final report, including an analysis of existing real-time evaluation tools in Annex, provided in hard copy (1 copy) and electronic version in Microsoft Word 97.
- Organised collection of existing real-time evaluation tools, in electronic format (Microsoft Word 97, Excel or PDF), or where this is not possible, in hard copy.
- A collection of key documentation obtained from other agencies (such as, examples of real-time evaluations, analysis of real-time evaluation experience) where these are not already available on the Internet, in electronic format (Microsoft Word 97 or PDF), or where this is not possible, in hard copy.
- A listing of all documents submitted with electronic file names should be submitted in hard copy and electronic format.

All electronic files should be presented on one CD.

Work arrangements
The consultant will be free to work independently from any base location, providing his/her own work equipment and space.

The total time required for the work involved is estimated to be one month, but pacing of work may be organised according to the consultant's preferred schedule.

Fees, expenses and payment arrangements
Airfare, DSA, and other expenses during the Briefing visit will be covered to the value of entitlement provided to UNICEF staff.

Reasonable costs for courier services, purchase of reference materials, etc. will be reimbursed in the final payment upon presentation of a list of expenses and the submission of the reference materials.

Payment is a lump sum of US$9,000 and will be made in two phases:
- One at a mid-point in carrying out data collection and analysis (date to be determined on finalising the contract) on submission of a brief report of progress (US$4,500);
- One on submission of all final products ($4,500).