

UNDERSTANDING FACTORS TO EVALUATION UPTAKE: A USER CENTERED APPROACH TO ADAPTING EVALUATIONS AND LEARNINGS

AUGUST 2021



USAID
FROM THE AMERICAN PEOPLE



INTERNATIONAL
RESCUE
COMMITTEE

Table of Contents

INTRODUCTION.....	2
METHODS AND ADAPTATIONS	3
Evaluation Methods.....	3
1. Emergency Response Review	3
2. After Action Review	3
3. Hybrid.....	4
Data collection tools to solicit user feedback	4
Iteration sessions	4
COVID-19 Adaptations	4
FINDINGS.....	5
Thematic Trends	5
ERR	6
AAR.....	7
Hybrid modalities.....	7
Conclusion.....	9

This study is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of International Rescue Committee, Inc. and do not necessarily reflect the views of USAID or the United States Government.

INTRODUCTION

Evaluations and learning exercises conducted in emergency responses allow for rapid course correction and accountability to individuals experiencing acute crisis through improved humanitarian responses. However, evaluations are only as successful as their findings, and recommendations are deemed useful and actionable by end users or decision makers. The International Rescue Committee (IRC) sought to understand enabling and limiting factors associated with the evaluation and learning process and its ensuing recommendations by implementing several evaluation methodologies across responses and soliciting feedback from the users on their experience. This approach was used to begin to unpack what motivates teams to conduct the exercises and act on their findings.

Real Time Evaluations¹ (RTEs) with modified methodologies have been used by the IRC since 2012 with the aim of producing an immediate snapshot of the strengths and challenges in an emergency response to empower urgent, corrective action. However, there are important challenges to conducting evaluations of emergency programming, including short timelines, over-burdened staff, security and safety concerns, de-prioritization of evaluation activities, and rapidly evolving contexts.² These challenges limit the conduct of evaluations but more importantly the uptake of evaluation findings. Even when data are available, it may not always be used to improve programming if the follow through is lacking. This means resources may be spent on evaluations in emergency settings which may or may not lead to improved interventions for emergency-affected populations.

The IRC implemented three different modalities of program evaluation ranging from an in-depth mixed methods approach (Hybrid approach including RTEs), a light-touch qualitative approach (After Action Review³ or AAR) to a purely quantitative scorecard (Emergency Response Review⁴ or ERR) over the last two years in fourteen different emergency responses, and sought to understand the characteristics that make the approaches useful as well as changes that would need to be made for future applicability. Because use of, and interest in participating in, evaluations or learning exercises requires a high-level of buy-in, a commitment was made to continuous, user-informed improvements. Reviews, termed iteration sessions, were periodically held over the course of the two years to review user feedback about the evaluation as well as facilitators and barriers to implementation and uptake.

While COVID-19 substantively changed the trajectory of the proposed study, an exploratory review is shared here to provide an overview of the process implemented to improve the methodologies and initial learnings gleaned from this process. A user-centric approach to exercise adaption was seen as an integral part of ensuring uptake when the exercise is not mandatory. The objective is to share this process and learning in the event it can inform quality improvement of evaluation and learning exercises for others in the humanitarian sector.

¹ <https://rescue.box.com/s/u3ynsg4rsd0tozuveusar2vh80yvztw9>

² <https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Real-time-evaluation.pdf>

³ <https://rescue.box.com/s/ur51rw8i4oi8qcnlzymc4rlbsz6fr3we>

⁴ <https://rescue.box.com/s/uc5f9k3vjc3sehadiciuxmmr81qt3tvm>

METHODS AND ADAPTATIONS

Evaluation Methods

In this section, an overview will be provided of the types of evaluation exercises conducted and reviewed. The different methods were employed as they each bring their own approach and style to learning and evaluation. All methods are expected to be conducted approximately three months into an emergency response. Additionally, details about adaptations to facilitate implementation of the evaluations and learning exercises in the COVID-19 context are included.

1. Emergency Response Review⁵

The Emergency Response Review (ERR) is an exercise developed to assess key performance indicators (KPIs) of a quality emergency response. The indicators capture program, operational, and response management criteria, that when taken together is designed to give the user a broad overview of key response outputs, activities, and decisions around planning and implementation through a stop-light scoring system. Indicators performing well are highlighted in green, while indicators in red are performing poorly, and indicators in yellow have room for improvement. Due to the use of KPIs and the final output being a scorecard, the ERR is categorized as a quantitative method. The goal of the ERR exercise is to:

- Identify if minimum standards are met
- Identify potential areas of concern
- Provide response teams with a general indication of where to investigate further
- If multiple exercises are conducted over time, provide an overview of change in performance

Data collection includes document reviews, staff surveys, and beneficiary surveys by a Monitoring, Evaluation, Accountability and Learning (MEAL) lead and if not available, a Response Coordinator internal to the team.

2. After Action Review⁶

The After Action Review (AAR) exercise is a half day facilitated discussion about response programming, operations and response management, open to all staff engaged in the response. The goal of the AAR exercise is to:

- Identify areas of success
- Identify potential areas of concern
- Identify what could be used for future responses

Like the Emergency Response Review, the After Action Review enable teams to discuss areas of success and concern. The key difference is that the After Action Review enables a semi-structured discussion about successes and concerns while the ERR requires users to focus on specific indicators. Facilitators are identified by the response manager or lead to conduct the exercise. Facilitation is ideally led by the Response Coordinator/Lead and MEAL Coordinator/Lead. Once the exercise is completed, the facilitators are responsible for collating the findings from the group exercises to produce a report. The AAR is characterized as a qualitative method due to the focus group style discussion that takes place.

⁵ Exercises completed in Uganda, Pakistan, South Sudan, and Niger

⁶ Exercises completed in Burundi, Chad, Bangladesh, Myanmar, Afghanistan, and Malaysia

3. Hybrid⁷

A third set of exercises characterized by the hybrid modalities of data collection were grouped together. Core sources of information include a combination of at least key informant interviews and focus group discussions. Real Time Evaluations also include staff and beneficiary surveys, which lead to the characterization of a mixed-methods approach. Data collection is completed over an approximately two-week period by a team of two to three staff and culminates in a report of findings and recommendations. The evaluators identified for these exercises are external to the response team to bring a neutral perspective. These exercises are typically implemented in larger scale responses due to the resourcing required to carry them out.

Data collection tools to solicit user feedback

The primary source of data that informed adaptive changes was through an initial reaction interview⁸ designed for use with response management. Interviews were conducted approximately one week after the completion of the learning or evaluation exercise. Questions for the interview were developed using the factors that affect utilization of evaluations as outlined by the Asian Development Bank⁹. Based on the feedback received across multiple exercises, the findings were coded and aggregated for thematic analysis to identify relevant changes to both tools and processes.

Iteration sessions

Once several learning exercises of the same kind were completed, iteration sessions were scheduled with the Emergency MEAL team. Data from initial reaction interviews were analyzed thematically and discussed to identify adaptations to learning tools or processes that would improve usability of the method. The approach was designed in such a way that any changes to the evaluation method was driven by user feedback with a focus on ease of use, clarity of process and improvement around uptake of recommendations. Once the required changes were identified through the iteration session, edits were made directly to the respective tools.

COVID-19 Adaptations

Prior to implementing the learning and evaluation methods with teams, the global COVID-19 Pandemic necessitated adaptations to the evaluation methods and the data collection tools for user feedback that would inform our iteration sessions – to reduce workload for teams having to rapidly adapt their existing programs to COVID as well as to keep compliant with COVID-19 public health and travel measures.

Three tools were initially developed to help us compare utilization and feasibility of these three learning methods: 1) the initial reaction interview outlined above, 2) a utilization survey, and 3) a cost and time tracker. The survey was designed to solicit expanded feedback against the utilization criteria, while the cost and time tracker was meant to help conduct a comparative feasibility analysis, examining cost, time and human resource requirements across the three learning methods. To minimize burden on response teams as they adapted programming to respond during the pandemic, the utilization survey had to be

⁷ Exercises completed in Sudan, Lebanon, and for the IRC Global Covid Leadership Team Response

⁸ <https://rescue.box.com/s/3rut1jsbgvrtw81pg5aqwab7ow8jun4b>

⁹ <https://www.adb.org/sites/default/files/evaluation-document/35880/files/evaluation-findings.pdf>

dropped. While the cost and time tracker was initially implemented, it was later dropped as the data collected was not yielding sufficient actionable insight to warrant continued use.

With regard to the learning methods, changes were made to minimize person to person contact. Focus group discussions with staff were modified so that input could be provided using a collaborative whiteboard via Mural¹⁰ and Microsoft Teams¹¹ for audio and video conferencing. Mural was selected as the online platform of choice as it enabled both anonymous and asynchronous input and was free. Microsoft Teams was recommended as the communication platform due to the broad organizational use of the tool as well as the ability for facilitators to use the breakout room functionality and solicit input from the response team in an additional way. Due to travel restrictions, methods requiring key informant interviews were done remotely rather than in person.

FINDINGS

Thematic Trends

Trends that appeared across exercise types included: 1. Ensuring availability of evaluation methods evaluators can use to answer a wider range of questions decision makers have. 2. The need for explicit engagement of response level senior management in the review of and accountability for recommendations. 3. Launching communication about the exercise and setting expectations closer to the start of the response to improve the timeliness of implementation of the exercise and 4. Overcoming staff availability and workload issues, a barrier to implementation, with additional co-facilitation support.

It is not uncommon at the culmination of an evaluation for some parties to express that the process did not address the questions they expected it would. This was commonly identified when a standard evaluation or learning method was used that was designed to answer a pre-defined set of questions. The mismatch between expectations and output translated to reduced confidence in findings and recommendations being less relevant for those who identified an issue with the scope. To remedy the mismatch, having a suite of adaptable evaluation or learning method tools that address a broad range of questions is important. By drawing from a suit of flexible options, evaluators are able to avoid using an evaluation methodology that was designed to answer a fixed set of questions.

Response managers play a critical role in the uptake of recommendations. Some of the managers that were interviewed indicated that expectations of reviewing findings, recommendations, and next steps were not clear when draft findings were shared with them. User recommendation to remedy this gap was that exercise findings and recommendations should be sent to the response senior management team for their input on recommendations as well as assigning responsible parties to move the recommendations forward before the exercise findings are finalized. The inclusion of pre-defined channels or responsible parties to review and address the uptake of recommendations was indicated as a measure to reinforce accountability.

Teams that indicated delays in implementation of the learning exercise cited multiple causal factors including staff turnover, staff bandwidth, and competing priorities. In certain instances, teams were unable to participate in the completion of an exercise due to one or more of the stated issues. While these

¹⁰ <https://www.mural.co/>

¹¹ <https://www.microsoft.com/en-us/microsoft-teams/group-chat-software>

issues tend to be pervasive within the sector, mitigative action can be taken by the evaluation team to potentially eliminate or reduce implementation delays. Feedback from interviewees indicated that engagement with teams at the onset of an emergency response about evaluation questions they would face and activities they may want to consider would enable teams to schedule necessary data collection activities ahead of time and better account for staff turnover.

For teams that identified that they did not have the capacity to implement the AAR exercise by themselves when the activity was being negotiated, additional facilitation support was provided by either regional and/or global colleagues. This served to demonstrate to the teams the resources required to successfully implement an activity, develop self-sufficiency for the team to implement in the future, and attempt to demonstrate the value-add of conducting an evaluation. In settings where exercises are not mandatory, supportive co-facilitation served to promote implementation and in turn increased the number of regional staff with experience in facilitating these exercises. This experience should presumably help support the overall quality and efficacy of the outputs.

ERR

Review of the ERR feedback revealed three trends or areas for improvement to support better implementation: 1. Indicators should be adaptable to the context of start-up of the response, including existing programming, stakeholders and coordination mechanisms. 2. Staff composition for data collection and completion of the exercise should include one MEAL focal point and one response coordinator (or a role that is aware of the relevant response activities and indicators) in order to expedite identification and review of relevant response documents. 3. Data collection activities for the ERR exercise should align with existing data collection efforts to minimize costs and staff time of additional M&E activities.

Key performance indicators that are not contextually flexible affect the perceived relevance of the indicator as well as any composite value in a scorecard that the indicator is a part of. Users of the exercise noted that some indicators were not applicable to their respective emergency response. The common trend among the indicators identified were that they assumed expected outputs in all start-up contexts rather than assessing for an output that is dependent on the context. Performance indicators should be developed under the assumption that there is flexibility to account for varying start-up scenarios including existing programming, stakeholders and coordination mechanisms.

Staffing changes throughout an emergency pose significant challenges to data collection if consistent communication and information management systems are not in place from the start. This is not always feasible. Feedback indicated that having only a MEAL focal point engaged in the data collection efforts slowed down completion of the exercise, as a MEAL focal point may not have been engaged in the development of specific response outputs or communications. A measure suggested to ensure relevant data could be tracked better was to ensure a two-person data collection team of a MEAL focal point and a response coordinator/lead. The two-person approach to data collection should reduce the time needed to complete the exercise as the response lead should have the ability to identify relevant documentation, while the expertise to deploy beneficiary and staff surveys in a timely manner would sit with the MEAL focal point.

During an emergency response, there are numerous competing priorities with insufficient resources to meet program objectives. Teams indicated that the flexibility to integrate exercise survey questions into

existing data collection efforts not only helped reduce workload and minimize costs, but also factored into their decision to conduct the exercise. When other data collection efforts are ongoing, teams should consider integrating the ERR beneficiary survey questions into existing surveys if it does not negatively delay discussing the findings of the exercise.

AAR

Findings from the AAR iteration sessions pointed towards the need to modify activities so that additional flexibility is available for teams to come together and focus on any concerns that they have about the response including areas for improvement and understanding the intersection of programs and operations. In order to achieve this level of flexibility, guidance was modified to ensure teams could select the activities that they could benefit the most from based on the questions they had about the response. Through this approach, response teams could discuss topics that were prioritized by them as opposed to using the initially proposed topics outlined within the AAR methodology.

Across the exercises reviewed, all interviewees indicated that the AAR exercise brought to light findings which were relevant to them. The uniform agreement of relevance of findings was only observed in the AAR modality. Feedback suggested that the opportunity to come together as a team to discuss all aspects of the response as well as the chance to discuss the intersection of all response components played a role in the relevance of the exercise.

Hybrid modalities

The key trends that emerged from the feedback after hybrid exercises were related to: 1. Perception of evaluation team composition/experience and 2. The process of negotiating the scope of the exercise.

Evaluation credibility is identified by the Asian Development Bank as a criterion for evaluation utilization. Feedback from interviewees indicated a perceived lack of familiarity with team practices and the response context as a factor that detracted from evaluation credibility. While the concern needs to be further examined for confirmation, a potential solution that was recommended to address it is through diversifying team composition to include evaluators that are from or have worked within the region or country of the emergency. Additionally, to address interviewee perceptions, two proposed solutions included sharing evaluator CVs with the response team and having evaluators prioritize interviews with senior response staff first. Sharing evaluator CVs will potentially highlight contextual and evaluation experiences of the evaluators to the response team. Prioritizing interviews with senior response staff at the onset of the evaluation process can enable evaluators to get up to speed about the broad emergency response and adapt subsequent interviews to ensure relevance to response team practices.

The second trend was that exercise scope seemed to impact the relevance of exercise. Feedback on the scope of the exercises were mixed such that some found the activity to be too broad while others found the exercise to not include key components. Further interrogation of this issue is required to understand if these findings are due to chance or if there is an exercise-specific or process issue that needs to be resolved. Interviewees provided tangible recommendations on how to address the issue; by defining who the target audience for the report is and tying that to clear goals of the exercise, consumers of the findings should be able to identify why certain topics are not covered. In scenarios where there are multiple goals, it is important to consider the alignment of the key goals with the data collection methods to ensure that findings are representative of the questions being asked and if not, evaluate the need to modify the

exercise in a way where data collection methods will provide answers to relevant questions and communicate the limitations of the approach with stakeholders so that expectations are managed.

Conclusion

In order to apply the findings of the review, evaluation teams should consider some of the following practices, if not already doing so:

It is important to discuss implementation of a future learning exercise or response evaluation at the start of an emergency response. The ideal entry point could vary and possible avenues for response and MEAL teams to discuss implementation of future exercises may include initial response coordination calls, initial funding requests, and when programming decisions are being made. Making this entry point standard would also facilitate future learning exercise discussions.

Once evaluation goals are identified and data collection methods are selected, a review of the approach should be considered to identify limitations. The identification of limitations can foster effective communication with decision makers about the scope of the exercise as well as serve as a call to action for modifications to the goals and data collection methods. Evaluators at this stage should also consider ensuring clarity around the target audience for the exercise so that the limitations can be further contextualized. The IRC Emergency MEAL team has found having a suite of different methodologies available to use helps close this gap between expectation and output.

Establishing explicit guidance on the process of finalizing evaluation recommendations, including soliciting feedback, from whom, and assigning recommendation accountabilities serves to mitigate the risk of recommendations not being acted upon.

For the IRC team, utilizing existing structures such as a response management team or senior management team to review and finalize recommendations within the exercise is instrumental to ensuring actionable priorities are specified.

With an increasing number of teams conducting evaluation and learning exercises within emergency responses, the user centered approach to improving usability of the resources available is serving as a foundation to a much broader process of continuous improvement. The practice of soliciting feedback about the exercise – including the barriers and value add, should continue and should be explored by teams if they are not already doing so. As response teams at the IRC continue to develop the capacity to conduct these exercises, decision-makers continue to provide input on what aspects of these exercises are most useful to them, and engagement with teams on the implementation of the various available tools, there is an opportunity to increase the utilization of evaluations and learning exercises.