HUMANITARIAN ACADEMY · HUMANITAIRES JOBS

Mini MBA Programmes and MEAL: design, manage and prove impact

Course handout · Week 9 : Evaluating: criteria, terms of reference, matrix and use

A programme of exactly 3 months (12 weeks, 144 hours) to design, manage and evaluate humanitarian programmes: needs assessment and severity, results chain and indicators, targeting and cash transfers, planning and budget, MEAL plan and data quality, sampling, analysis, accountability, evaluation, impact measurement, value for money and adaptive management. Corrected numerical exercises, three monthly exams with case studies, and a verifiable diploma from the Humanitarian Academy.

https://humanitairejobs.org/en/academy/mini-mba-programmes-meal/semaine-9/

Free document from the Humanitarian Academy. It may be printed and shared freely for training purposes, without modification.

Week 9 : Evaluating: criteria, terms of reference, matrix and use

Objectives

- Choose the type of evaluation and the criteria suited to the decision it must inform.
- Write precise terms of reference and build a rigorous evaluation matrix.
- Manage an evaluation independently and turn its recommendations into tracked decisions.

Why and when to evaluate

An **evaluation** is a systematic and objective assessment of an intervention, its design, its implementation and its results. It differs from monitoring: monitoring is continuous and internal, and asks "are we doing what we planned?"; evaluation is periodic, often external, and asks "are we doing the right thing, in the right way, and what difference does it make?".

It has two functions, sometimes in tension: **accountability** (reporting to people, donors and the board) and **learning** (improving this programme and future ones). An evaluation designed only for accountability produces scores; an evaluation designed for learning produces decisions.

Type	Timing	Main question
Real-time evaluation	First weeks of an emergency	Should the response be corrected now?
Mid-term evaluation	Middle of the project	What needs adjusting for the second half?
Final evaluation	End of the project	What results, what lessons for the future?
Impact evaluation	During and after	What change is attributable to the intervention?
Ex-post evaluation	1 to 3 years after the end	Have the effects lasted?

Before commissioning an evaluation, write down who will use it and for what decision. If no one can answer, the evaluation will be one more report on a shelf.

The OECD-DAC criteria

The evaluation criteria of the OECD Development Assistance Committee, revised in 2019, are six in number:

- **Relevance**: does the intervention respond to people's needs and priorities, and does it remain relevant if the context changes?
- **Coherence** (added in 2019): is it compatible with other interventions in the same context, both internal and external to the organisation?

- **Effectiveness**: does it achieve its objectives, including for different groups?
- **Efficiency**: does it deliver its results economically and on time?
- **Impact**: what longer-term effects, positive or negative, intended or not?
- **Sustainability**: will the benefits continue after funding ends?

In humanitarian work, the criteria adapted by ALNAP are also used: **coverage** (have all people in need been reached, without exclusion?), **connectedness** (link between the emergency and the longer term) and **coordination**.

The criteria do not all carry the same weight. A real-time evaluation of a three-month response focuses on relevance, coverage and effectiveness; talking about sustainability at that stage makes little sense. Choosing three or four criteria and going into them in depth is better than skimming over six.

Terms of reference that get a good evaluation

The **terms of reference** (ToR) are the intellectual contract of the evaluation. Vague ToR produce vague evaluations. They contain:

- **Context** and description of the intervention (with the theory of change and the logframe in annex).
- **Purpose and users**: why now, for whom, for which decisions.
- **Scope**: period, areas, components, what is excluded.
- **Evaluation questions**: five to eight main questions, not thirty. Each question can be evaluated with the means and time available.
- **Expected approach and methods**, leaving the team to make proposals, and the existing data available.
- **Ethics**: consent, data protection, protocols for sensitive topics, participation of people.
- **Deliverables and timeline**: inception report, debriefing, draft report, final report, summary.
- **Team profile**, budget, governance (reference group) and independence rules.

Order of magnitude. A good-quality external final evaluation of a multisectoral project often costs between 25,000 and 80,000 USD depending on the field setting and methods. Too small a budget leads to a few days of fieldwork and a report without evidence; fewer evaluations, better done, is preferable.

The evaluation matrix and the inception report

The **evaluation matrix** is the central tool of the inception report: for each question, it specifies the sub-questions, the indicators or judgement criteria, the data sources, the data collection and analysis methods, and the limitations.

Example. Question: "To what extent have cash transfers improved the food security of the most vulnerable households?" Judgement criteria: change in the food consumption score and rCSI between baseline and endline, compared with a comparison area; differences according to

the sex of the head of household and the presence of a person with a disability. Sources: baseline and endline surveys, post-distribution monitoring, focus group discussions. Methods: difference-in-differences, qualitative analysis by group. Limitations: comparison area partially assisted by other actors from month 6.

The **inception report** is the moment to correct course: the team confirms what is feasible, flags questions that cannot be addressed, and proposes the sampling. The evaluation manager validates it with the reference group before fieldwork. It is the best investment of time in the whole evaluation.

The **quality** of an evaluation is judged on the clarity of the link between **findings** (supported by data), **conclusions** (judgements based on the findings) and **recommendations** (actions that flow from the conclusions). A recommendation without a corresponding finding is the evaluator's opinion.

Independence, ethics and evaluation management

Independence protects credibility: the evaluation team took no part in the design or implementation; the evaluation manager is not the head of the project being evaluated; the team has free access to data and people; the programme may comment on the facts in the draft report, but may not impose changes to the conclusions.

The United Nations Evaluation Group (UNEG) norms and standards for evaluation and the ALNAP principles insist on **ethics**: do no harm to the people interviewed, informed consent, confidentiality, attention to people in distress and to risks of retaliation, feedback to communities.

The **evaluation manager** has an active role: preparing data, organising field access, convening the reference group, checking quality at key stages, protecting the team's independence, and organising the debriefing. The main risks are well known: start-up delays that shorten fieldwork, non-existent baseline data, limited access, pressure to soften findings.

> An evaluation that finds only positive points is not credible. A programme that accepts well-supported negative findings shows that it is serious.

Management response and use

The **management response** is the document that turns an evaluation into decisions. For each recommendation: accepted, partially accepted or rejected, with the justification; planned actions; person responsible; deadline. It is written within a month of the final report and published with it.

Follow-up of actions takes place at quarterly management reviews: where does each action stand? Mature organisations publish an annual implementation rate for accepted recommendations.

A recommendation can legitimately be rejected: too costly, outside the mandate, already under way. But the justification must be written down. What is not acceptable is the absence of a response.

Finally, **dissemination** adapts the format to the audience: a two-page summary for management and donors, a presentation for teams, a simple oral version for communities, and publication of the full report (for example in the ALNAP library) for the sector. Publishing evaluations, including critical ones, is a Grand Bargain commitment on transparency.

Exercise: writing the questions and matrix of an evaluation

Situation. An 18-month project (water, hygiene and protection) in three displacement camps reaches its midpoint. Country management must decide within four months whether and how to extend the project to a fourth camp. The donor has requested an evaluation but has not set any questions. Budget: 30,000 USD. Baseline data exist for water, not for protection.

Task. 1) What type of evaluation and which main criteria? 2) Write three evaluation questions. 3) For one of them, fill in one row of the matrix (judgement criteria, sources, methods, limitations).

Answer. A formative mid-term evaluation, with country management and the donor as users; main criteria: relevance, effectiveness, coverage, efficiency (sustainability is secondary at this stage). Possible questions: 1) To what extent do the water and hygiene services meet the planned standards for all groups, including people with disabilities? 2) Do the protection activities respond to the priority risks identified by women and children? 3) What conditions must be met to extend the project to a fourth camp with the same quality and a comparable cost per person? Matrix row for 2): criteria: match between reported priority risks and activities, use of services, perception of safety; sources: focus group discussions separated by sex and age, interviews with protection actors, service data (anonymised); methods: coded qualitative analysis, participatory risk mapping; limitations: no protection baseline, hence a contribution analysis rather than a measurement of change.

Key takeaways

- Monitoring asks "are we doing what we planned?"; evaluation asks "is it the right thing, done well, and what difference does it make?".
- Six DAC criteria (2019): relevance, coherence, effectiveness, efficiency, impact, sustainability; plus coverage, connectedness and coordination in humanitarian work.
- Precise ToR: users, decisions, five to eight evaluable questions, ethics, deliverables.
- The evaluation matrix links each question to criteria, sources, methods and limitations.
- Management response: each recommendation accepted or not, with action, person responsible and deadline.

Quiz

1. Which criterion was added by the OECD-DAC in 2019?
   A. Relevance
   B. Coherence
   C. Efficiency
   D. Impact

2. An evaluation that asks whether all people in need were reached, without exclusion, addresses the criterion of:
   A. Sustainability
   B. Coverage
   C. Efficiency
   D. Connectedness

3. The most useful moment to correct a poorly focused evaluation is:
   A. Reading the final report
   B. Validating the inception report and the evaluation matrix
   C. Publication
   D. The management response

4. A recommendation that is not based on any finding in the report is:
   A. Good practice
   B. The evaluator's opinion, to be challenged
   C. Mandatory
   D. A conclusion

5. Who should not manage the evaluation of a project?
   A. A country MEAL manager
   B. The head of the project being evaluated
   C. A headquarters adviser
   D. A steering committee

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B