



GLOBAL EVALUATION REPORT OVERSIGHT SYSTEM

Handbook for UNICEF Staff & Independent Assessors

unicef 
Evaluation Office

GEROS Handbook

Version 3.2

Cover photo © UNICEF/ UN019061



June 2017

UNICEF GEROS Handbook

- This Handbook aims to orient UNICEF staff and independent assessors to the GEROS tool and to the GEROS system more generally.
 - This is part of a wider commitment to aid better understanding of UNICEF’s evaluation function and to support accountability.
- The Handbook includes guidance for independent assessors on providing clear and comprehensive feedback on evaluation reports; enabling evaluation commissioners, senior managers and the executive board to make decisions based on a clear understanding of the quality of evaluation evidence.
 - The handbook helps to set a common standard and level of expectation between UNICEF staff and the independent GEROS assessment team about the GEROS process and the evaluation quality assessment
- Improving evaluation report quality is intended to contribute to:
 - Enhancing the usefulness of evaluations
 - Building UNICEF corporate knowledge and organizational learning
 - Strengthening evaluation capacity for better evaluation in the future.
- Evaluators may also find the Handbook to be useful as a clear articulation of UNICEF expectations regarding a good evaluation report.
 - Preparing evaluation reports that meet UNICEF standards can help ensure completeness, reduce the need for revisions and edits, and improve uptake of recommendations.

Contents

- Chapter 1: Evaluation Quality Assurance
 - An overview of what evaluation quality assurance is and why it is important to UNICEF
- Chapter 2: GEROS Assessment Process
 - How the GEROS system works and what can be expected
- Chapter 3: Assessment Standards
 - Detailed information on each element of the evaluation quality assessment
- Chapter 4: Review Tools
 - Links to the key GEROS tools and further resources

June 2017

How to use this handbook

For UNICEF staff

- UNICEF staff may be most interested in
 - Chapter 1 (an introduction to evaluation quality assessment)
 - Chapter 2 (the GEROS assessment process)
- Other chapters are very detailed and need only be referenced to help understand a specific question or issue
 - Chapter 3 (assessment standards)
 - Chapter 4 (tools)

For independent assessors

- Assessors should be familiar with Chapters 2, 3 and 4 in detail
 - Chapter 1 helps to better understand the context of GEROS

June 2017

Acronyms

CEE/CIS	Central and Eastern Europe and Commonwealth of Independent States [Region]
CCC	Core Commitments to Children
CEDAW	Convention on the Elimination of all forms of Discrimination Against Women
CO	Country Office
CRC	Convention on the Rights of the Child
EAPR	East Asia and Pacific Region
ERDB	Evaluation and Research Database
EO	Evaluation Office
EPI	Evaluation Performance Indicator (for UN-SWAP)
EQA	Evaluation Quality Assurance
ESAR	Eastern and Southern Africa Region
GEEW	Gender Equality and Empowerment of Women
GEROS	Global Evaluation Report Oversight System
HQ	Headquarters [New York]
HRBAP	Human Rights Based Approach to Programming
LACR	Latin America and the Caribbean Region
MENA	Middle East and North Africa [Region]
PPP	UNICEF Programme Planning Process
OECD-DAC	Organisation for Economic Cooperation and Development – Development Assistance Committee
RBM	Results Based Management
RO	Regional Office
ROSA	Regional Office for South Asia
SPOA	Strategic Plan Outcome Area
TOR	Terms of Reference
UNEG	UN Evaluation Group
UN-SWAP	UN System Wide Action Plan for gender equality
WCAR	West and Central Africa Region

June 2017

*An overview of what evaluation quality assurance is and why it is
important to UNICEF*

CHAPTER I: EVALUATION QUALITY ASSESSMENT

The GEROS system

- The purpose of GEROS is to support strengthening of the evaluation function to meet and exceed United Nations Evaluation Group (UNEG) standards, UN System Wide Action Plan on gender equality (UN-SWAP) and other UNICEF commitments (including equity and human-rights based approaches).
 - GEROS aims to ensure accountability and to promote the use of robust evaluative evidence.
- The GEROS system consists of systematic and independent quality assessment of evaluation reports that have been uploaded to the corporate Evaluation and Research Database (ERDB) by UNICEF Country Offices, Regional Offices, HQ divisions and Evaluation Office.
 - Assessments of individual reports are fed back to commissioning offices and publicly available along with the evaluation reports online.
 - Annual meta analysis reports identify global, regional and thematic trends in the quality of evaluation reports.
- GEROS has three objectives (revised in 2016)
 - Objective 1: Enabling environment for senior managers and executive board to make informed decisions based on a clear understanding of the quality of evaluation evidence and usefulness of evaluation reports
 - Objective 2: Feedback leads to stronger evaluation capacity of UNICEF and partners
 - Objective 3: UNICEF and partners are more knowledgeable about what works, where and for who.

June 2017

Evaluation quality assessment

- The evaluation function, guided by the evaluation policy and PPP, seeks to strategically contribute to UNICEF performance by providing good-quality evidence for:
 - decision making
 - policy advocacy and strategic positioning
 - organizational learning and accountability.
- Reflecting the decentralized nature of UNICEF, the majority of evaluations supported by UNICEF are managed at the decentralized level.
 - The decentralized nature of the evaluation function ensures that evidence generated is relevant to the local, national and regional context and therefore more likely to serve purposes that include informing national policies for children.
 - However, all decentralized evaluation functions pose the challenge of setting up a consistent corporate system to ensure good quality and credibility.
 - The Global Evaluation Reports Oversight System was designed to help address this challenge.
- Evaluation quality assessment is therefore important to:
 - Quantifying the quality of evaluation reports to help inform UNICEF managers (CO, RO, HQ) and partners on one aspect of the current performance of the evaluation function
 - Provide an incentive for evaluation managers, assessors and evaluators to make all efforts to improve the quality of evaluation reports before being submitted to GEROS
 - Provide constructive written feedback to help improve the credibility and use of future evaluations, thus contributing to the professionalization of the evaluation function over time
 - Identify trends in quality over time to inform evaluation strategies and policies
 - Help to identify good evaluation reports to be included in evaluation syntheses

June 2017

The GEROS principles

- GEROS is an organization-wide system under the Evaluation Policy.
- The EQA (evaluation quality assessment) of final evaluation reports is managed by the Evaluation Office (EO).
- GEROS is complemented by, and independent from, a number of regional quality-assurance mechanisms, which provide assessments of draft ToR and draft reports against the UNICEF/UNEG ToR and Report checklists, giving real time feedback to allow quality improvement of final ToR and reports.
- Some UNICEF offices have established evaluation quality assurance mechanisms that can refer to the GEROS tools and standards.
- GEROS is informed by the UNEG Norms and Standards for Evaluation 2016
- UNICEF-adapted UNEG standards are the basis for quality assessment of final evaluation reports.
- GEROS integrates the UN System Wide Action Plan for gender equality and empowerment of women (UN-SWAP) Evaluation Performance Indicator (EPI).
- To ensure credibility and objectivity, the quality assessments of final evaluation reports is undertaken by an external and independent firm selected through an open bidding process.
- While the corporate Evaluation and Research Database (ERDB) contains evaluations, surveys, studies and researches, GEROS focuses on the assessments of evaluation reports.
- Evaluations are defined by using PPP Manual definitions and the UNICEF evaluation office Taxonomy.
- EO screens reports uploaded to the ERDB, and submits the reports identified as “evaluations” to the external firm for a final screening and evaluation quality assessment.
- EO may reclassify reports as evaluations when they have been misclassified

June 2017

The GEROS quality rating scale

- Since 2016, GEROS uses a 5 point scale based upon the level to which reports meet assessment criteria
- UN System Wide Action Plan for gender equality (UN-SWAP) criteria are rated according to the methods set by the UN Evaluation Group; with the results integrated into the GEROS rating according to the following table.

Section & Overall Rating	UN-SWAP Rating	Implication	GEROS Score
Highly Satisfactory	Exceeds Requirements (3)	Exceeds UNICEF/UNEG standards for evaluation reports. Decision makers may use the evaluation with a high degree of confidence	4
Satisfactory	Meets Requirements (2)	Meets UNICEF/UNEG standards for evaluation reports. Decision makers may use the evaluation with confidence	3
Fair	Approaches Requirements (1)	Meets UNICEF/UNEG standards for evaluation reports in some regards, but not all. Decision makers may use the evaluation with caution. Substantive improvements in some areas are needed	2
Unsatisfactory	Misses Requirements (0)	Does not sufficiently meet the UNICEF/UNEG standards for evaluation reports. Decision makers cannot rely on the evaluation.	1
Missing		Important aspects of the evaluation that are required by the UNICEF/UNEG standards were found to be absent. The evaluation report is incomplete.	0
Not Rated	Not Applicable	Refers to aspect of the evaluation that was not rated for a legitimate reason. Normally this does not affect the quality of evaluation report.	0

June 2017

How the GEROS system works and what can be expected

11

CHAPTER 2: GEROS ASSESSMENT PROCESS

Standards for a good quality evaluation report

- An evaluation report is assessed as being of good quality when:
 - It is a credible, independent and useful report that addresses the evaluation purpose and objectives.
 - Can be used with confidence by decision-makers.
- Evaluation reports are assessed using the UNICEF-adapted UNEG evaluation report standards to assess the following core elements:
 - Clear and full description of the background of the evaluation
 - The evaluation's purpose, objectives and scope are fully explained
 - Appropriate and sound methodology in line with the ethical standards and procedure
 - Findings, conclusions and recommendations are based on evidence, sound analysis and are useful
 - Lessons learned are correctly identified.
 - Well structured, logical and clear report
 - Meet or exceed UN-SWAP evaluation performance indicator for integration of gender equality

June 2017

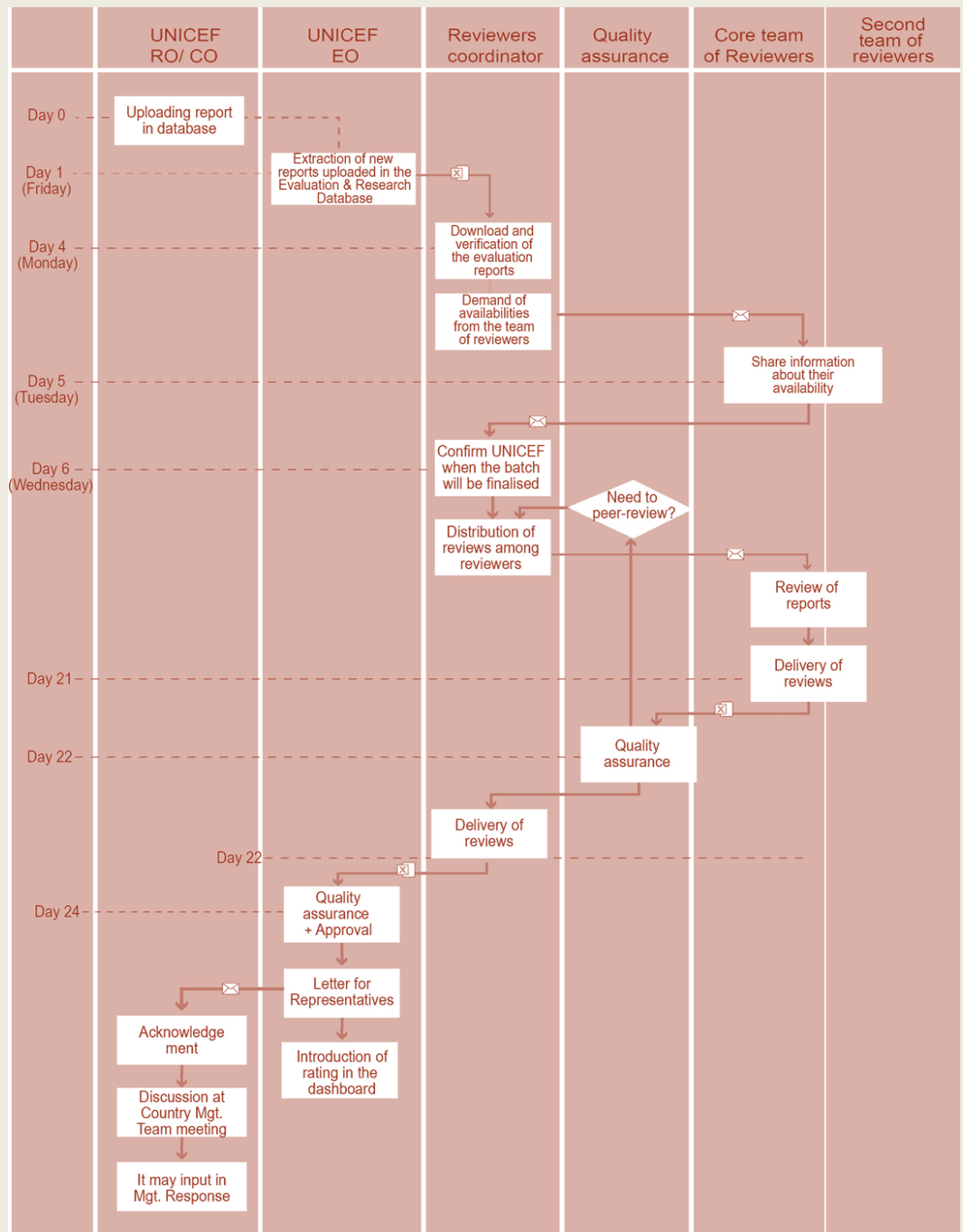
Standards for a good quality GEROS assessment

- A good quality GEROS assessment (evaluation quality assessment) is:
 - Timely, submitted within the agreed timeframe
 - Complete and in the original language of the evaluation report
 - Accurate and reliable, giving examples where relevant. Does not second-guess the evaluation process itself.
 - Consistent between quantitative and qualitative ratings, and across the evaluation portfolio
 - Constructive in tone and proportionate in the application of standards to the scale and nature of the evaluation being assessed
 - In accordance with the ‘spirit and purpose’ of the GEROS standards. An important consideration in the process that criteria emphasize the quality of elements, not just the extent to which that element is present
 - Honest, independent and quality controlled

June 2017

Workflow for the assessment process

The workflow for GEROS starts with the uploading of evaluation documents to the ERDB by the section in charge of the evaluation.



June 2017

Approach to the EQA assessment (1)

- Every assessment begins with a thorough reading of the evaluation report, to orient the assessor to the report and to give an overview of the various components. Following this, the assessor should:
 - Go through the report in detail, using the GEROS tool’s section by section approach
 - Concentrate on responding to the particular questions for each section by reading the report and analyzing the extent to which it satisfies each of the criteria.
- Once the ratings for the criteria have been completed (and thus the questions given their automatic ratings), the assessor should determine an overall ‘section’ score that is coherent with the question scores (‘section’ scores are automatically weighted and used to calculate the overall score by the EQA template).
 - This is an iterative process, with the assessor using the determination of the section rating as an opportunity to reexamine their criteria scores and to adjust these if necessary.
 - A key consideration is ensuring consistency between the overall section rating and the question ratings.
- The next stage is to write the executive feedback for each section.
 - The assessor should pick up on particularly strong criteria and comment on what makes them strong by either explicitly or implicitly drawing upon evaluation standards and practices, as well as picking up on particularly weak criteria and commenting on why they are weak, again by either explicitly or implicitly drawing upon evaluation standards and practices.

June 2017

Approach to the EQA assessment (2)

- Following written feedback on each section, the assessor should:
 - Work on the “additional comments for section X”, where they should pick up on the weak points highlighted in the “executive feedback” section and should concretely explain how the evaluation report could have been improved as regards that section.
 - Comments might also focus upon the section overall rather than particular criteria. In so doing, reference should be made to evaluation standards and practices.
- UNICEF is particularly interested in section H, “Evaluation Principles”.
 - Assessing this section requires reading through the report and seeing that there is not just a passing reference made to key issues such as ethics, gender equality, equity and human rights-based approach, but that they are actually integrated into the report.
- The final stage is an overall scan through the assessment, to determine if criteria and section ratings are appropriate, as well as if there is good consistency between these.
 - Comments should be reviewed and adjusted as necessary to strengthen them.
 - At the end, the assessor should complete the section “Lessons for Improving the Management and Performance of Future Evaluations”. Here, the assessor can concentrate on elaborating lessons learned that emerge from the assessment.

Weighting of quality assessment ratings

To ensure consistency of ratings while preserving the freedom for professional assessment by the assessors, the overall rating for an evaluation report is based on a weighted aggregation of the scores from eight EQA 'sections'.

The weightings are set and followed-up at the Annual Review Workshops for GEROS.

Each section is given a quality score (0-4) by an assessor based on a set of guiding questions. Each question has several indicators that are first scored 0-4. The overall section score should be guided by the average ratings of the indicators for each question in that section, but this is not automatic and the assessor maintains professional discrepancy.

The justification for the score is included in written feedback by the assessor; and the EQA tool displays the average scores for each question to support transparency.

Section	Weighting (%)
Background	5
Purpose	5
Methods	15
Findings	20
Conclusions / Lessons	15
Recommendations	15
Structure	5
Principles	15
Executive Summary	5
TOTAL	100

June 2017

Contents of each assessment

- Background information (meta data)
 - Title of the Evaluation Report; Report sequence number; Region; Year of Report; Office; Countries covered; ToRs present; Date of Review; Name of assessor
- Classification of Evaluation Report (UNICEF taxonomy)
 - Management of Evaluation (Managerial control and oversight of evaluation decisions); Strategic Plan Outcome Areas (SPOA) Correspondence (Alignment with SPOA focus area priorities); Evaluation object; Evaluation type; Evaluation strategy; Evaluation design; Evaluation level; Geographic Scope
- Main assessment
 - Sections A-I; Overall Rating; Feedback on Evaluation Management
- Executive summary (auto-generated)
 - Overall rating; Section ratings; Executive feedback
- Data sheet (auto-generated)
 - All data from the single row for inserting in database for analysis

June 2017

Style of qualitative comments

The 'house style' for GEROS comments is **constructive and clear**

GEROS assessments are prepared in the **original language** of the evaluation report (Ar, En, Es, Fr, Po, Ru)

Narrative is aligned with the ratings for consistency. Comments should use words and structure sentences that avoid dissonance between the comments and the ratings. If a section rates as "fair" then comments should not use a term like "very good" even if the context it is used in is accurate.

June 2017

- **Be a peer, not the police:** use words like "could", "may consider", is "advised to" and avoid instructional words such as "should" unless this is warranted by the complete absence of a UNICEF standard (such as included the ToR as an annex).
- **Be a guide to the UNEG standards:** use phrases such as "the UNICEF/UNEG standards require that..." instead of saying that something is missing or is needed – help the reader to hold themselves to account.
- **Acknowledge constraints and flag things that are contentious:** where the evaluation has clearly faced constraints because of the context, let the reader know that these are acknowledged and understood in the comments, even if the standards require the report to be rated low.
- **Use constructive feedback:** start the comment by stating the indisputable facts. Then explain the implications of those facts on evaluation report quality (with references). Finally, provide guidance on how the issue can be addressed next time. If possible, refer back to a positive example in the same report that can be built on, or give an example of a specific solution to similar challenges in the future.
- **Go beyond the indicators – feedback must add value:** avoid repeating the information from the indicators within the text. Try to elaborate underlying patterns or why the evaluation may have faced difficulties – for example if the recommendations are vague because the purpose was unclear.
- **Try to be specific:** where possible, refer to examples from the report that support your comment. Where there are gaps, share specific ideas for the future (such as "future evaluations of this nature may wish to consider cluster sampling"). Be familiar with the full range of approaches and techniques on BetterEvaluation.org and all UNEG guidance

EQA reports and follow-up

EQA report

- The EQA process generates 5 products for use by UNICEF staff
 1. A full copy of each independent assessment (Excel and PDF)
 2. An executive summary of each independent assessment (PDF)
 3. A letter to the senior manager responsible for each evaluation explaining the EQA result
 4. The UNICEF Evaluation Dashboard
 5. An annual meta analysis report with feedback for each UNICEF region

Management response

- Management responses should be uploaded within 60 days.
 - Management responses are independent of the EQA process and should not rely on the availability of a completed EQA
 - Management responses are required for all evaluations. Evaluation reports rated as Unsatisfactory by GEROS still require a management response. In these cases, the MR can identify specific findings that will be followed-up or triangulated with other data.

*Detailed information on each element of the evaluation quality
assessment*

CHAPTER 3: ASSESSMENT STANDARDS

Meta-data (used for analysing overall trends)

Assessment details

- Title of the Evaluation Report
- Report sequence number
- Region
- Year of Report
- Office
- Coverage (countries)
- ToRs present
- Date of Review
- Name of reviewer

June 2017

Classification of the evaluation

- Management of Evaluation
- SPOA Correspondence
- Evaluation object
- Evaluation level
- Geographic Scope
- UNICEF Taxonomy
 - Evaluation type
 - Evaluation strategy
 - Evaluation design

Assessment sections, questions and indicators

- ▶ SECTION A
- ▶ SECTION B
- SECTION C
- SECTION D
- SECTION E
- SECTION F
- SECTION G
- SECTION H
- SECTION I

June 2017

SECTION A: BACKGROUND (weight 5%)

Question 1. Is the object of the evaluation clearly described?

Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status

Clear and relevant description of intended beneficiaries by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cities, sub-regions...) and in terms of numbers reached (as appropriate to the purpose of the evaluation)

Description of the relative importance of the object to UNICEF (e.g. in terms of size, influence, or positioning)

Question 2. Is the context of the intervention clearly described?

Clear and relevant description of the context of the intervention (policy, socio-economic, political, institutional, international factors relevant to the implementation of the intervention)

Clear and relevant description (where appropriate) of the status and needs of the target groups for the intervention

Explanation of how the context relates to the implementation of the intervention

Question 3. Is the results chain or logic well articulated?

Clear and complete description of the intervention's intended results

Intervention logic presented as a coherent theory of change, logic chain or logic framework

Question 4. Are key stakeholders and their contributions clearly identified?

Identification of implementing agency(ies), development partners, primary duty bearers, secondary duty bearers, and rights holders

Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF

SECTION B: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)

Question 5. Is the purpose of the evaluation clearly described?

Specific identification of how the evaluation is intended to be used and to what this use is expected to achieve
Identification of appropriate primary intended users of the evaluation

Question 6. Are the objectives and scope of the evaluation clear and realistic?

Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR

Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention)

Assessment sections, questions and indicators

SECTION A

SECTION B

▶ SECTION C

SECTION D

SECTION E

SECTION F

SECTION G

SECTION H

SECTION I

June 2017

SECTION C: EVALUATION METHODOLOGY (weight 15%)

Question 7. Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation?
UNICEF evaluation standards refer to the OECD/DAC criteria. Not all OECD/DAC criteria are relevant to all evaluation objectives and scopes. Standard OECD DAC Criteria include: Relevance; Effectiveness; Efficiency; Sustainability; Impact. Evaluations should also consider equity, gender and human rights (these can be mainstreamed into other criteria). Humanitarian evaluations should consider Coverage; Connectedness; Coordination; Protection; Security.

Clear and relevant presentation of the evaluation framework including clear evaluation questions used to guide the evaluation

If the framework is OTHER than UNICEF standard criteria, or if not all standard criteria of the chosen framework are included, the reasons for this are clearly explained and the chosen framework is clearly described

Question 8. Does the report specify methods for data collection, analysis, and sampling?

Clear and complete description of a relevant design and set of methods that are suitable for the evaluation's purpose, objectives and scope

Clear and complete description of the data sources, rationale for their selection and sampling strategy. This should include a description of how diverse perspectives are captured (or if not, provide reasons for this), how accuracy is ensured, and the extent to which data limitations are mitigated

Clear and complete description of the methods of analysis, including triangulation of multiple lines and levels of evidence (if relevant)?

Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias?

Question 9. Are ethical issues and considerations described?
The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:

Explicit reference to the obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability)

Description of ethical safeguards for participants appropriate for the issues described (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm)

ONLY FOR THOSE CASES WHERE THE EVALUATION INVOLVES INTERVIEWING CHILDREN: explicit reference is made to the UNICEF procedures for Ethical Research Involving Children

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

▶ SECTION D

▶ SECTION E

SECTION F

SECTION G

SECTION H

SECTION I

June 2017

SECTION D: EVALUATION FINDINGS (weight 20%)

Question 10.	Do the findings clearly address all evaluation objectives and scope?
	Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions and criteria
	Reference to the intervention's results framework in the formulation of the findings
Question 11.	Are evaluation findings derived from the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence.
	The evaluation clearly presents multiple lines (including multiple time series) and levels (output, outcome, and appropriate disaggregation) of credible evidence.
	Findings are clearly supported by and respond to the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison.
	Unexpected effects (positive and negative) are identified and analysed
	The causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or- not- from implementation to results).
Question 12.	Does the evaluation assess and use the intervention's Results Based Management elements?
	Clear and comprehensive assessment of the intervention's monitoring system (including completeness and appropriateness of results/performance framework -including vertical and horizontal logic; M&E tools and their usage)
	Clear and complete assessment of the use of monitoring data in decision making

SECTION E: EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 15%)

Question 13.	Do the conclusions present an objective overall assessment of the intervention?
	Clear and complete description of the strengths and weaknesses of the intervention that adds insight and analysis beyond the findings
	Description of the foreseeable implications of the findings for the future of the intervention (if formative evaluation or if the implementation is expected to continue or have additional phase)
	The conclusions are derived appropriately from findings
Question 14.	Are lessons learned correctly identified?
	Correctly identified lessons that stem logically from the findings, presents an analysis of how they can be applied to different contexts and/or different sectors, and takes into account evidential limitations such as generalizing from single point observations.

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

SECTION E

▶ SECTION F

▶ SECTION G

SECTION H

▶ SECTION I

June 2017

SECTION F: RECOMMENDATIONS (weight 15%)

Question 15. Are recommendations well grounded in the evaluation?

Recommendations are logically derived from the findings and/or conclusions

Recommendations are useful to primary intended users and uses (relevant to the intervention and provide realistic description of how they can be made operational in the context of the evaluation)

Clear description of the process for developing recommendations, including a relevant explanation if the level of participation of stakeholders at this stage is not in proportion with the level of participation in the intervention and/or in the conduct of the evaluation

Question 16. Are recommendations clearly presented?

Clear identification of target group for action for each recommendation (or clearly clustered group of recommendations)

Clear prioritization and/or classification of recommendations to support use

SECTION G: EVALUATION STRUCTURE/PRESENTATION (weight 5%)

Question 17. Does the evaluation report include all relevant information?

Opening pages include:

Name of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents -including, as relevant, tables, graphs, figures, annexes-; list of acronyms/abbreviations, page numbers

Annexes should include, when not present in the body of the report:

Terms of Reference, Evaluation matrix, list of interviewees, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence

Other appropriate annexes could include: additional details on methodology, copy of the results chain, information about the evaluator(s)

Question 18. Is the report logically structured?

The structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles)

Context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations

SECTION I: EXECUTIVE SUMMARY (weight 5%)

Question 22. Can the executive summary inform decision-making?

An executive summary is provided that is of relevant conciseness and depth for primary intended users

Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key findings, key conclusions, key recommendations)

Includes all the necessary information to understand the intervention and the evaluation AND does not contain information not already included in the rest of the report

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

SECTION E

SECTION F

SECTION G

▶ SECTION H

SECTION I

June 2017

SECTION H: EVALUATION PRINCIPLES (weight 15%)

Question 19.	Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity?
	Reference and use of rights-based framework, and/or CRC, and/or CCC, and/or CEDAW and/or other rights related benchmarks in the design of the evaluation
	Clear description of the level of participation of key stakeholders in the conduct of the evaluation, and description of the rationale for the chosen level of participation (for example, a reference group is established, stakeholders are involved as informants or in data gathering)
	Stylistic evidence of the inclusion of these considerations can include: using human-rights language; gender-sensitive and child-sensitive writing; disaggregating data by gender, age and disability groups; disaggregating data by socially excluded groups.
Question 20.	Does the evaluation assess the extent to which the implementation of the intervention addressed gender, equity & child rights?
	Identification and assessment of the presence or absence of equity considerations in the design and implementation of the intervention
	Identification and assessment of the presence or absence of gender in the design and implementation of the intervention
	Explicit analysis of the involvement in the object of right holders, duty bearers, and socially marginalized groups, and the differential benefits received by different groups of children
	Clear proportionality between the level of participation in the intervention and in the evaluation, or clear explanation of deviation from this principle (this may be related to specifications of the ToRs, inaccessibility of stakeholders at the time of the evaluation, budgetary constraints, etc.)
Question 21.	Does the evaluation meet UN SWAP evaluation performance indicators?
	Note: this question will be rated according to UN SWAP standards
	GEEW is integrated in the Evaluation Scope of analysis and Indicators are designed in a way that ensures GEEW-related data will be collected
	Evaluation Criteria and Evaluation Questions specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.
	A gender-responsive Evaluation Methodology, Methods and tools, and Data Analysis Techniques are selected.
	The evaluation Findings, Conclusions and Recommendation reflect a gender analysis

Meta data

- Assessment details
- Evaluation classification

Used for the meta analysis

Meta Data

ELEMENTS	CONTENT
Title of the Evaluation Report	The title as it appears on the cover of the report.
Report sequence number	UNICEF's internal report number, in the format year of report submission-report number name of the country. E.g.: "2016-002 Ethiopia - Evaluation".
Region	UNICEF region: <ul style="list-style-type: none"> · CEE/CIS · EAPR · ESAR · MENA · ROSA · WCAR · LACR · HQ
Year of Report	Year when the report was submitted to the Evaluation Reports Data Base (ERDB). It should coincide with the Report's sequence number
Office	The UNICEF office that commissioned the evaluation
Coverage	Country/countries the evaluation is focusing upon.
ToRs present	The Terms of Reference (ToRs) for the evaluation are present in the evaluation report. Yes/No
Date of Review	Date the report was assessed
Name of assessor	<ul style="list-style-type: none"> · ImpactReady · Independent

June 2017

Evaluation taxonomy (I)

Management of Evaluation (Managerial control and oversight of evaluation decisions)	Who managed the evaluation, according to the evaluation report or the ToR: <ul style="list-style-type: none"> · UNICEF managed · Jointly managed with one or more UN agencies · Jointly managed with organizations outside the UN system · Jointly Managed with Country · Country-led (government) Evaluation · Externally managed (e.g. by donors) · Not clear from Report 	
SPOA Correspondence (Alignment with SPOA focus area priorities)	<ul style="list-style-type: none"> • Health • HIV/AIDS • WASH • Nutrition • Education • Child protection • Social inclusion • Gender equality (cross-cutting) • Humanitarian action (cross-cutting) 	
Evaluation object	<ul style="list-style-type: none"> · Pilot/innovation · Project · Programme · Country Programme · Joint Programme · Organization/business unit · Policy/Norms/Standards · Thematic area (e.g. health, sanitation, violence against children) · Strategy · System 	
Evaluation type	<ul style="list-style-type: none"> · Formative 	Mid-term evaluation with learning purpose
	<ul style="list-style-type: none"> · Summative 	Final evaluation mainly done for accountability
	<ul style="list-style-type: none"> · Summative & formative 	Final evaluation for learning and accountability
	<ul style="list-style-type: none"> · Meta Evaluation 	Evaluation or review of an evaluation

June 2017

Evaluation strategy	· Quantitative	Evidence comes from surveys and numeric data
	· Qualitative	Evidence comes from interviews and focus groups and descriptive data
	· Mixed methods	The evaluation design relies on both types of data to build the evidence of the findings
Evaluation design	Experimental	An evaluation design with an experimental group and control group in which the researcher introduces an intervention and measures its impact on the dependent variable at least two times
	Quasi-experimental	A research/evaluation design in which participants are not randomly assigned to treatment conditions, but in which comparison groups are constructed by statistical means.
	Case study	This research design entails an in-depth examination of a single or several case(s) (e.g. individuals, groups, institutions, countries, processes), conducted and designed to result in a thorough and well- organized understanding of the subject(s) being examined.
	Comparative	A design which uses the comparison of two or more cases in order to illuminate existing knowledge or generate new insights as a result of contrasting of the findings
	Theory-based	A design in which the theory of an intervention is elaborated, and the data is collected to find evidence that supports or refutes this theory in practice.
	Retrospective	A research design based on the analysis of existing data (e.g., birth and death certificates, medical records, school records, or employment records) or by obtaining information about past events elicited through interviews or survey-questionnaires.
	Cohort/panel	A research design in which participants in a well-defined cohort, usually a group of individuals born in the same period of time, are followed over time.
	Participatory research	Action- A multi-stage participatory and inclusive research design, in which a problem is diagnosed collaboratively with the concerned stakeholders with the purpose of finding practical solutions to the problem.
	Systematic review	A research design that provides a framework for drawing together and critically assessing the results from existing evidence on a focused question or topic. The evidence-base is selected according to clear criteria and the review is conducted through a standardized protocol.
	Cross-sectional	This is a research design in which data from particular participants are obtained at a single point in time in order to collect a body of quantitative or quantifiable information in connection with two or more variables, which are then examined to detect patterns of association.
	Longitudinal	A research design in which data are collected from the same sample at different intervals at least two different times.

June 2017

Evaluation taxonomy (3)

Evaluation level	Output	Causal effects deriving directly from programme activities, and assumed to be completely under programme control
	Outcome	Effects from one or more programmes being implemented by multiple actors (UNICEF and others), where the cumulative effect of outputs elicits results beyond the control of any one agency or programme
	Impact	Final results of a programme or policy on the intended beneficiaries and, where possible, on comparison groups. Reflects the cumulative effect of donor supported programmes of cooperation and national policy initiatives.
Geographic Scope	National Multi-country Regional Multi-region/Global Other	

Section A: Background

Weighting = 5%

- Question 1. Is the object of the evaluation clearly described?
- Question 2. Is the context of the intervention clearly described?
- Question 3. Is the results chain or logic well-articulated?
- Question 4. Are key stakeholders and their contributions clearly identified?



Read an example of a comprehensive description of the object being evaluated from Ethiopia (Ethiopia 2016/012)

June 2017

- Question I. Is the object of the evaluation clearly described?
 - Element I.1. Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status.
- Detail on whether the intervention is a project, programme, strategy, humanitarian response, etc.
- Locations and maps of where the intervention takes place,
- Timelines explaining the evolution of the intervention and implementation status (mid/end cycle)
- Overall budget and some yearly break-down
- Detail on the activities (number of target people, locations, duration), etc.

Yes	To be found in the report or in the annexes all these elements.
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the intervention is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 1.2. Clear and relevant description of intended recipients by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighborhoods, town/cities, sub-regions...) and in terms of numbers reached (as appropriate to the purpose of the evaluation).
- The target groups of the program, including break-down per component or per location (if more than one and relevant),
- Describe and quantify individuals, communities or organizations
- Reason why they were selected as target group
- How they are organized if they interact
- Difference between directly or indirectly benefited.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the beneficiaries is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

Example: Target groups and geographical coverage

From report: “Midterm evaluation report ‘Improving water, sanitation and hygiene (WASH) facilities in prioritized schools in the West Bank and Gaza Strip, 2012-2014’” UNICEF WASH in School Programme Alpha International for Research, Polling and Informatics, Ramallah.

August 2014
(Palestine 2015/002)

June 2017

“UNICEF addresses the WASH needs of a total of 207 schools (89 schools in the West Bank and

118 schools in Gaza) selected jointly with MoEHE based on a prioritized list of 246 schools.

Based on 2011 enrolment figures, the average number of students per school is 350 students and 950 students in West Bank and Gaza Strip respectively. As such, the estimated number of direct beneficiaries is 143,250 students (81,250 boys and 62,000 girls).

The targeted schools in the period 2012-2014 are:	WB	Gaza	Total
Year			
2012	29	39	68
2013	30	39	69
2014	30	40	70
Total	89	118	207

The selection of high priority schools requiring rehabilitation/construction of WASH facilities was done by the MoEHE at the district level Directorate offices in coordination with the School Health and School Building Services sections based on criteria including:

- Lack of toilets (students use toilets at nearby schools)
- Number of toilets below national standards of student to toilet ratio of 30:1
- Lack of separate toilets for boys and girls; and male and female teachers
- Dilapidated toilets that are non-functioning or poorly functioning
- Safety concerns where there is a risk of toilet collapsing as a result of large water tank on the top of the toilet block or toilets built over cesspits
- School expansions to include additional grade levels
- Schools changed from boys to girls or mixed
- Schools with disabled children

While boys’ schools are generally worse off in terms of WASH facilities, the selection of targeted schools by MoEHE took into consideration the need to include as many girls and co-education schools as possible to address the issue of girl dropping out of secondary schools due to lack of separate sanitation facilities for girls.”

- Element 1.3. Description of the relative importance of the object to UNICEF (e.g. in terms of size, influence or positioning).
- Intervention fits within the overall agenda of the Country Programme
- Within the Regional Office agenda
- Within the global agenda.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the importance is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).



Not to be mistaken with the context of the evaluation:

- The context of the intervention refers to the external factors that defined the situation and conditioned the intervention.
- The context of the evaluation refers to the internal (and external) factors that defined the situation at the beginning of the evaluation.

- Question 2. Is the context of the intervention clearly described?
 - Element 2.1. Clear and relevant description of the context of the intervention (policy, socio-economic, political, institutional, international factors relevant to the implementation of the intervention).
 - Political situation
 - Socio-economic environment
 - The Institutional context
 - Policy environment: National priorities, policies and plans in the sector, international commitments described.
 - International and national legal framework/s, the country's National Strategies or Plans, related SDG
 - Other national or international factors that may affect the intervention.
 - Data sources are indexed and relevant

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the context is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

Evaluation indépendante de la réponse
humanitaire de l'UNICEF à la crise nutritionnelle
en Mauritanie entre 2012 et 2015
Evaluation menée entre février et avril 2016



RAPPORT D'ÉVALUATION

Version Finale octobre 2016

Elaboré par une Equipe de Bac-Santé

- Dr Mohamed Nethrou HAMED, Consultant principal, expert en santé publique
- Dr José María ECHEVERRÍA HERNÁNDEZ, Consultant associé, expert en action humanitaire
- Dr Tidiane NDOYE, Consultant associé, expert en recherche qualitative



Evaluation indépendante de la réponse humanitaire à la crise nutritionnelle en Mauritanie 2012-2015 Page 1

Read an example of a comprehensive description of the context from Mauritania (Mauritania 2016/022)

- Element 2.2. Clear and relevant description (where appropriate) of the status and needs of the target groups for the intervention.
- Reason why the target groups' status or needs justifies the intervention
- Demographics
- Access to services for different gender/ ethnicities
- Detail about the situation of groups described in Element 1.2.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the needs of the target groups is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Example

From report: Evaluation of the
Thailand National Child and Youth
Development Plan 2012-2015
(Thailand 2015/003)

“There are approximately 800,000 births each year in Thailand. The estimated total child and youth population (0-25 years) was 22.2 million in 2013, of which 11.4 million male and 10.8 female, compared to 23.5 million in 2008 (11.8 million male and 11.7 female).

Thailand is on track to achieve most of the Millennium Development Goals (MDGs) in 2015. Poverty levels have continued to fall: the number of people living below the national poverty line decreased from 28.4% in 1992 to 8.1% in 2009 with a short period of increase due to the financial crisis of 1997.

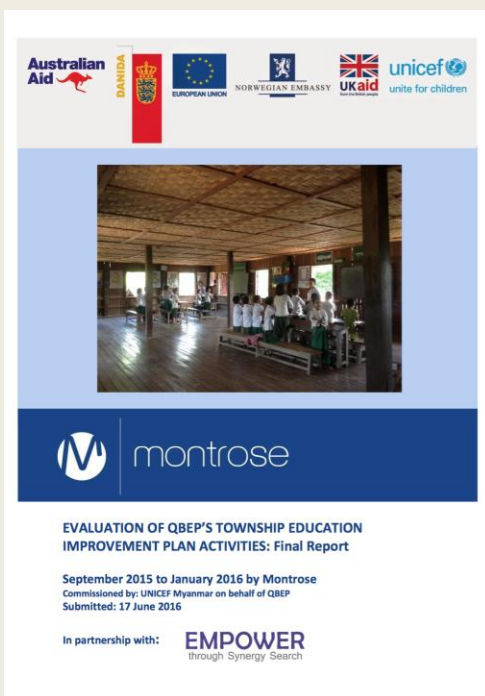
The wellbeing of children and women in Thailand has continued to improve. There has been significant progress in terms of basic health care provided to children and regarding their wellbeing. Thailand has considerably reduced the numbers of child mortality in recent decades. Since 1990, achievements included a 50% fall in infant mortality with immunization coverage above 90%. Thailand’s Under-5 mortality rate during 2005–2010 was 15 per 1000 live births, and decreased to 13 per 1000 live births in 2013.

However, analysis of MICS data from 2005/6 and 2012 shows a worsening nutrition status at national level for children of 0-59 months of age. There is indeed a significant increase in moderate and severe underweight as well as moderate wasting and severe stunting; at the same time, there is a significant increase in moderate overweight. Low levels of breastfeeding, with only 11.2% of infants exclusively breastfed up to the age of 6 months and only 37.3% of children predominantly breastfed, appear to play an important role in this respect. These figures are among the lowest in the Southeast-Asia region.”

- Blankenship, Jessica, *Comparison of MICS3 and MICS4 datasets for anthropometric status of children 0-59 months of age: Analysis of variables influencing children’s nutrition status in Thailand. Summary Version.* November 2014.

- **Element 2.3. Explanation of how the context relates to the implementation of the intervention**
 - How needs are addressed by the intervention
 - Rationale for choosing the selected strategies
 - Other relevant factors in the context regarding the implementation
 - Data sources are listed and relevant

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the context is provided
Not Rated	For some reason, this element is not applicable (Reason explained).



Read an example of a reconstructed theory of change from Myanmar (Myanmar 2016/013)

- Question 3. Is the results chain or logic well-articulated?
- Element 3.1. Clear and complete description of the intervention's intended results
- The report very clearly presents the interventions' intended goal, outcome and output results
- Initial needs analysis or process that led to designing the intervention as it is.
- Potential Baseline information

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the intended results is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

Example: listing the main results statements from a Programme

Final evaluation report: “Developing Community based Services for Children with Disabilities and their Families” (April 2010 –December 2013). (Serbia 2013/010)

June 2017

Overall and Specific Objectives

The overall objective of the Project “Developing community based services for children with disabilities and their families” is to strengthen policy frameworks, institutional mechanisms and capacities so that community-based services supporting children with disabilities can be developed and are responsive to the needs of these children and their families.

The specific objective of the Project is to increase the number of children with disabilities that are benefiting from community services that are clearly contributing to their social inclusion.

Purpose and Expected Results

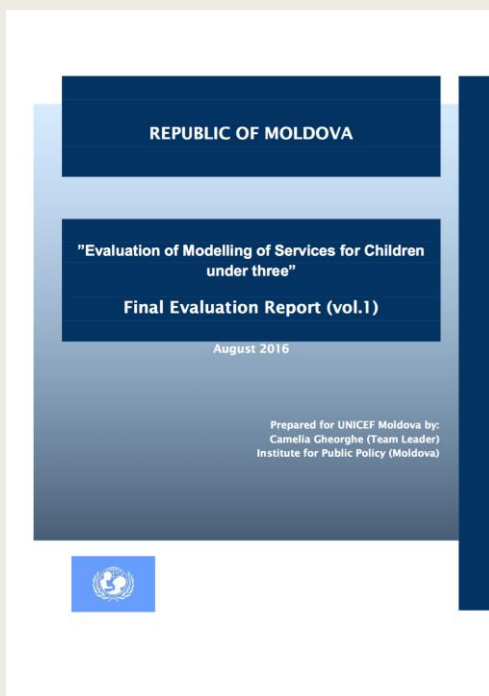
According to the **Theory of Change**, the main purpose of the Project is to increase the number of children with disabilities who are benefiting from community services that are clearly contributing to their social inclusion, with two main expected results:

1. Ministry of Labour, Employment and Social Policy capacity strengthened to monitor, evaluate and supervise decentralized and well-targeted community based social protection services in the Republic of Serbia (national level)
2. Municipal authorities, service providers, centres for social work and civil society activists capacitated to fulfil their respective roles in ensuring community services for families with children with disability are accessible and meet set standards and procedural guidelines (local level).

The main project **outcomes** target, as per **Theory of Change**, a wide range of national and local government institutions as well as interested civil society representatives, with an overall aim to provide effective CBSS for children with disability and their families. The Theory of Change further develops the two overarching **results** into a number of core **activities** for each implementation stage: three groups of activities at the national level (1.1 to 1.3) and four at the municipal level (2.1 to 2.4).

- Element 3.2. Intervention logic presented as a coherent theory of change, logic chain or logic framework
 - Intended results are presented
 - Logic between activities and results is clear and discussed
 - The rationale (why choosing those strategies) for the intervention design is justified
 - The mechanisms the actions are expected to cause so that results are achieved are also made explicit.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the theory of change is provided
Not Rated	For some reason, this element is not applicable (Reason explained).



Read an example of explaining the role of UNICEF from Moldova (Moldova 2016/001)

June 2017

- Question 4. Are key stakeholders and their contributions clearly identified?
- Element 4.1. Identification of implementing agency (ies), development partners, primary duty bearers, secondary duty bearers, and rights holders
- A descriptive or visual representations (stakeholders mapping) of all the actors involved in the intervention: implementers, partners, target groups, and their types and scale.
- The stakeholders are classified or described according to their roles, in terms of primary duty bearers, secondary duty bearers, right holders, etc.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about stakeholders involved is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Example: key stakeholders

From: Evaluation of the Thailand
National Child and Youth
Development Plan 2012-2015
(Thailand 2015/003)

The National Commission on the Promotion of Child and Youth Development (NCPCYD), which was established based on the NCYDPA, is the highest authority on child and youth development issues and a key stakeholder of the NCYDP. Eight line ministries also played an important role in the development and implementation of the plan including the following: Ministry of Social Development and Human Security (MSDHS), Ministry of Education, Ministry of Public Health, Ministry of Labour, Ministry of Justice, Ministry of Culture, Ministry of Interior, and Ministry of Information and Communication Technology.

Other key stakeholders in the development and implementation of the plan at national level were the National Statistics Office, civil society organizations and Universities. UNICEF also provided support to the plan's implementation.

At sub-national level, the provincial Governor's office, the Provincial Social Development and Human Security Office, and the departments of the above-mentioned ministries all played an important role as well as the Provincial Committee on Child Protection, established under the National Child Protection Act, as they had taken up, to varying degrees, some aspects of child and youth development. At local level, key stakeholders included the Tambon Administrative Office and its social welfare and education office. Moreover, Child and Youth Councils at national, provincial, district and Tambon level proved important players with regards to child and youth participation.

The contributions of the various national and sub-national agencies consist of the portions of their regular budgets allocated to child and youth issues. These could not be quantified in the process of the evaluation.

The implementation of the plan depended on the activities the various national and sub-national agencies had included in it. Consequently, while the plan covered the period 2012-2016, the implementation of several of its measures had actually started earlier and continued during the period of the plan. Further details about the agencies involved and the roles they played are provided as part of Annex 4.

- Element 4.2. Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.
- The report clearly displays how the different groups or stakeholders (included UNICEF) were supposed to contribute to the intervention
- How that role was accomplished during the implementation.
- Differences between financial and other kinds of contributions.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the stakeholders' contribution is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Example

Final report of RuSFAD Phase III
(Pakistan 2013/003).

The project involved multiple stakeholders, both in the public and civil society sectors. The key public agencies engaged in project delivery included Ministry of Climate Change (MoCC) at the federal level, provincial and district public health engineering, health, education, local government and disaster management departments. The role of MoCC was to ensure coordination with UNICEF and provide policy guidance to keep programme consistent with PATS. The provincial governments and Public Health Engineering Department (PHED) were kept involved throughout the project cycle to demonstrate success of joint working. The UNHABITAT contributed to initial design and worked in the earlier phases, however the agency had disengaged from the project by the time third phase had started. UNICEF mobilized the major funding, provided technical backstopping and remained involved in project oversight.

Plan Pakistan was the lead of the consortium and tasked to implement the project through local Implementing Partners (IPs including NGOs /RSPs) in KP, Sindh, and Punjab. Plan Pakistan was responsible for financial management of resources to implement the project. Plan Australia provided technical backstopping to Plan Pakistan in areas such as sanitation marketing, designing IEC materials for mass media campaigns and conducting research and advocacy. All the IPs reported to Plan Pakistan.

WaterAid Pakistan joined in Phase II and continued as Third Party Monitors through Phase III. Their role was to develop a complete process and output driven M&E system, undertaken third party monitoring and provide performance assessment reports for management decision making. Knowledge management was another key function with which WAP was tasked.

June 2017

Section B: Purpose, Objectives and Scope

Weighting = 5%

- Question 5. Is the purpose of the evaluation clearly described?
- Question 6. Are the objectives and scope of the evaluation clear and realistic?

Tip!

The purpose relates to the one stated in the ToR, further developed and clarified between evaluators and commissioners during the inception phase.

- Question 5. Is the purpose of the evaluation clearly described?
- Element 5.1. Specific identification of how the evaluation is intended to be used and to what this use is expected to achieve
- What will be done with the evaluation outputs (findings, conclusions, recommendations, report) once the process has finalized.
- Documents or processes the report will be used as an input

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the use is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Example

From report: Evaluation report on
the Govt of Oman- UNICEF Child
Friendly School Initiative

(Oman 2016/001)

“The purpose of the evaluation is to inform the scaling-up potential of the CFS initiative to include all schools in Oman. Adopting a utilization-focused approach, this evaluation offers practical information to help the key stakeholders understand the extent to which the CFS initiative has positively influenced pilot schools (in terms of the learning environments and the learning outcomes for the children in these schools) and communities (in terms of level of engagement between the school and the community) in Oman, as compared to reference schools, since its inception in 2012.”

June 2017

EXAMPLE

This evaluation is aimed primarily at the humanitarian coordinator and the humanitarian country team, who are expected to use the results to ensure accountability and learning for the ongoing response; and the Inter-Agency Standing Committee principals, Working Group and Emergency Directors Group, who are expected to use Inter-Agency Humanitarian Evaluation results and lessons learned to contribute to global policy and practice.

From report: Inter-Agency Humanitarian Evaluation of the Response to the Crisis in the Central African Republic (HQ 2016/009)

- Element 5.2. Identification of appropriate primary intended users of the evaluation
 - Identification of the primary users of the evaluation, at local, regional and global level, within UNICEF and the rest of partners.
 - Identification of the secondary users.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the users is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

EXAMPLE

The main objectives of the evaluation were to:

- Appraise progress towards planned results.
- Develop a better understanding of the process of change.
- Draw any lessons that may help improve the implementation of the project.

The mid-term evaluation will inform the completion of the UNICEF/DFAT agreement and provide an opportunity for adjustment as required.

From report: “WATER, SANITATION & HYGIENE IN SCHOOLS PROGRAMME MID-TERM EVALUATION REPORT”. STATE OF PALESTINE AUGUST 2014 (Palestine 2015/002)

June 2017

- Question 6. Are the objectives and scope of the evaluation clear and realistic?
- Element 6.1. Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR.
- Expected outputs of the evaluation: what the commissioners sought to get out of the process.
- Reference to what is detailed in the ToR and justifies and discusses potential changes.
- The objectives relate to the ones stated in the ToR or explain their evolution.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the objectives is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Example of scope of evaluation

From the report: Evaluation of the Thailand National Child and Youth Development Plan 2012-2015

(Thailand 2015/003)

The evaluation covered the design of the intervention as well as its implementation during the period from 2012 until early 2015. It examined the four strategies underpinning the plan: increasing life immunity in children and youth; protection and development of children in need of special protection measures; capacity building of alliances for child and youth development; and improvement of the system for child and youth protection and development.

The evaluation assessed the added value of the plan with regards to the following: enhanced attention given to children and youth issues, better coordination among stakeholders, improved management of initiatives, and increased resource allocation. It also covered a selection of local-level projects as case studies in order to highlight good practices and identify enabling and constraining factors in the implementation of the plan's measures.

The evaluation examined the implementation of the plan at scale in the entire country. At national level, using data gathered from government and non-governmental stakeholders, the evaluation focused on national policies and programmes that address child and youth development, including early childhood development. At sub-national level, the evaluation focused on four selected provinces, located in four out of the five main regions in Thailand, i.e. Bangkok, Central, Northern, North-eastern and Southern regions, and two selected LAOs in each of province.

EXAMPLE

SCOPE OF THE EVALUATION

Geographical scope and Time period – the evaluation was designed to cover the period 2010-2015. The proposed approach entailed evaluating the intervention nationwide. The sample for both quantitative and qualitative data collection in this evaluation are *Abia, Adamawa, Anambra, Bauchi, Bayelsa, Edo, Katsina, Kebbi, Kwara, Niger, Ogun, Osun*), and *FCT Abuja*. Sampling approach and the limitations of the proposed design are presented in the companion report.

From Nigeria Evaluation of MNCH Week (Nigeria 2016/038)

June 2017

- Element 6.2. Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as the reasons for this scope (e.g., specifications by the TORs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).
- What will and will not be covered (thematically, chronologically, geographically with key terms defined),
- Reasons for this scope.
- It also discusses if that is the scope included in the ToR and reasons for any changes regarding it

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the scope is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Section C: Framework and Methods

Weighting = 15%

- Question 7. "Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation?"
- Question 8. Does the report specify methods for data collection, analysis, and sampling?"
- Question 9. "Are ethical issues and considerations described?"

EVALUATION CRITERIA AND QUESTIONS

The final set of evaluation questions (EQ) addressed via this evaluation is presented below.

Table 3: MNCHW IE, evaluation criteria and key evaluation questions

Evaluation Criteria	Key Evaluation Questions
Relevance	<ul style="list-style-type: none"> To what extent are the objectives of the programme still valid? Are the activities and outputs of the programme consistent with the intended impact? To what extent does the intervention reach its targeted clientele addressing its needs and priorities? To what extent is UNICEF support relevant?
Effectiveness and efficiency	<ul style="list-style-type: none"> To which extent has the MNCHW contributed to improve coverage of PHC interventions? To what extent has UNICEF and her partners support being effective and efficient? What are the intervention costs? Are there areas along the MNCHW delivery chain where operational efficiency improvements can be achieved? If so, what are they and how can they be implemented?
Impact	<ul style="list-style-type: none"> What is the consistency of the MCHW impact with the anticipated theory of change? What is the contribution of the MNCHW to sustain maternal, newborn and child health outcomes? What is the contribution of the MNCHW to strengthen the primary health care system?
Partnership	<ul style="list-style-type: none"> To what extent is the partnership between UNICEF and NPHCDA effective and coordinated? Are there clear roles and responsibilities of MNCH week partners?
Sustainability	<ul style="list-style-type: none"> How sustainable is the MNCHW? Is there an exit strategy in place? What are the roles of partners, and what is the level of ownership and accountability?

Example from Nigeria Evaluation of MNCH Week (Nigeria 2016/038)

June 2017

- Question 7. "Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation?"
 - **Element 7.1. UNICEF evaluation standards refer to the OECD/DAC criteria.** Not all OECD/DAC criteria are relevant to all evaluation objectives and scopes. Standard OECD DAC Criteria include: Relevance; Effectiveness; Efficiency; Sustainability; Impact. Evaluations should also consider equity, gender and human rights (these can be mainstreamed into other criteria). Humanitarian evaluations should consider Coverage; Coherence; Coordination; Protection.
- Each of the relevant criteria (usually Relevance, Effectiveness, Efficiency, Sustainability, and Impact) are described
- Including how they are understood in the evaluation context.
- Equity, gender and HR are also included as criteria, as well as other such as Coverage, Coordination, Protection, etc.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the evaluation criteria is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 7.2. Clear and relevant presentation of the evaluation framework including clear evaluation questions used to guide the evaluation.
 - The report clearly states the evaluation criteria
 - Guidelines and principles that will guide it.
 - Evaluation questions
 - Data sources
 - Analysis methods
 - The report includes an Evaluation Matrix gathering all this information

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the evaluation framework is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 7.3: If the framework is **OTHER** than UNICEF standard criteria, or if not all standard criteria of the chosen framework are included, the reasons for this are clearly explained and the chosen framework is clearly described.
- The report does a good job giving reasons documenting the changes regarding the ToR, most of all when the framework differs from UNICEF’s standard one

If the 5 standard criteria are not considered, then this element should read as “Not rated”.



If the evaluation is considering the DAC criteria but leaving Impact out of the scope, the report should mention this and the reason why, then the element would be rated as “Yes”.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about other criteria beyond the DAC set is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Tip!

Some reports include detailed methodological annexes

EXAMPLE

In Impact evaluations, the design is often clearly stated, but every evaluation (and every evaluation report) should discuss and define its design.

Examples of Design can be: Experimental, Quasi-experimental, Case study, Comparative before-after, Participatory, Action-research
Systematic review, Cross-sectional, Longitudinal, Cohort/panel, Retrospective, Theory-based.

- Question 8. Does the report specify methods for data collection, analysis, and sampling?
 - **Element 8.1. Clear and complete description of a relevant design and set of methods that are suitable for the evaluation's purpose, objectives and scope**
- The design (or how the evaluators will find reliable data to answer the questions)
- Methods are clearly described
- Mention on the mix qualitative-quantitative
- Rationale for the choices made (above other potential alternatives)
- Reference is made to the evaluation questions, the objectives, or the scope

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the design and methods is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

- Element 8.2. Clear and complete description of the data sources, rationale for their selection and sampling strategy. This should include a description of how diverse perspectives are captured (or if not, provide reasons for this), how accuracy is ensured, and the extent to which data limitations are mitigated.
- Who/what are the data sources
- Explanation on the selection and sampling strategies and size
- How data sources cover the diversity of actors.
- It also covers the acknowledgment and mitigation actions of the limitations.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the data sources and sampling is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Examples

From Report: Evaluation of the
Thailand National Child and Youth
Development Plan 2012-2015
(Thailand 2015/003)

June 2017

SAMPLE SIZE AND SAMPLING

We powered our study based on the ability to detect a significant change in one--day roll call absence (primary outcome), self-reported absence in the past seven days, and diarrhea recall in the past seven days. The mean risk at baseline and coefficient of variation was estimated using baseline data.

The original intention of the impact evaluation was to use the sub--sample of schools that had provided additional pupil information during the 2011 baseline. However, the conflict in the north made 32 of these schools inaccessible and lost to follow--up, and 16 of the remaining schools did not meet the criteria for a complete intervention package. Ultimately 42 of the beneficiary schools and 9 control schools that participated in the baseline were retained to participate in the impact evaluation.

LIMITATIONS TO THE METHODOLOGY

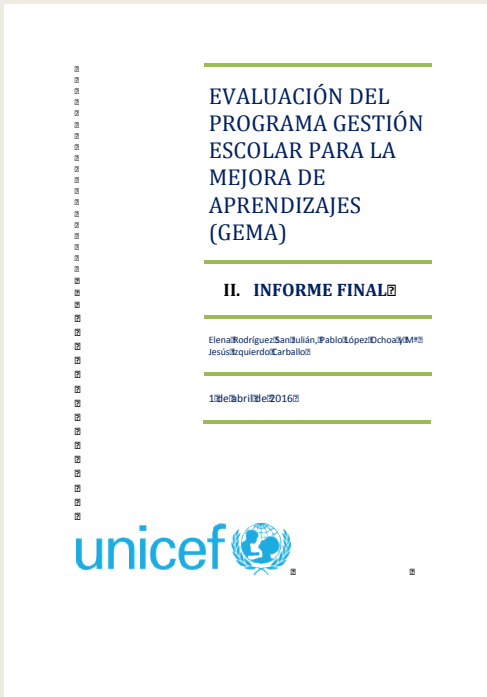
The number of provinces selected for field visits is relatively small number compared to the total number of 76 provinces in Thailand (the four selected provinces only represent 5.3% of all provinces). This is also true for the LAOs of which only two per province were selected. Thus, the total number of 8 selected LAOs represent only 0.1% of all 7,408 LAOs in the country (Census 2000), including Bangkok. This limited the countrywide representation of findings. The evaluation team mitigated against this risk by selecting each province in a different region, thus covering four out of the five regions of Thailand. Moreover, the findings of the field visits were complemented with interviews with stakeholders at the national level, which provided additional insight on the differences between provinces in terms of child and youth development.

- Element 8.3. Clear and complete description of the methods of analysis, including triangulation of multiple lines and levels of evidence (if relevant)?
 - Comprehensively covers how the evaluators analyzed the data
 - Explains how they assured that the data is reliable
 - Describes strategies used to triangulate findings (e.g. from different levels of evidence).

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the analysis is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 8.4. Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias?
 - Potential inconsistencies or gaps
 - Limitations of the data, the constraints faced by the evaluation
 - Gaps of data that the evaluators have faced during the analysis
 - Potential bias and the mitigation measures.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the limitations is provided
Not Rated	For some reason, this element is not applicable (Reason explained).



Read an example of an ethics section from Argentina (Argentina 2016/003)

June 2017

- Question 9. "Are ethical issues and considerations described?"
- *The evaluation should be guided by the UNEG ethical standards for evaluation.*
- Element 9.1. Explicit reference to the obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability)
- Clear reference to the principles the evaluators committed to, such as independence, impartiality, credibility, conflicts of interest, accountability, etc.
- Report discusses their implications.
- It discusses how potential conflicts (with the evaluation questions, scope, resources, report) and other conflicts of interest between commissioners and evaluators were prevented or resolved along the process

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the principles is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- **Element 9.2. Description of ethical safeguards for participants appropriate for the issues described (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm).**
 - Ethics safeguards are explained and ensured for participants
 - Respect for dignity and diversity
 - Right to self-determination
 - Fair representation
 - Compliance with codes for vulnerable groups
 - Confidentiality
 - Avoidance of harm.
 - Others

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the ethical safeguards is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Example

From the report: Evaluation of the
Thailand National Child and Youth
Development Plan 2012-2015

(Thailand 2015/003)

The evaluation team abided by the ethical code of conduct for UNEG evaluations as well as the UNEG Standards and Norms for Evaluation, in particular with regards to the independence of the evaluators, the respect of individual participants' anonymity and confidentiality, sensitivity to the social and cultural context as well as the evaluators' integrity and honesty in their relations with all stakeholders.

Moreover, in all phases of the evaluation process, the evaluation team abided by the ethical principles related to the evaluation of initiatives on violence against women (VAW) and gender-based violence (GBV) where applicable.

June 2017

- Element 9.3. For those cases where the evaluation involves interviewing children: Explicit reference is made to the UNICEF procedures for Ethical Research Involving Children.

- Specific procedures followed when interviewing children
- Ethical safeguards on their confidentiality, informed consent
- Appropriate language use

If children were not interviewed, this element should be “Not rated”.

Tip!

If children participated in interviews, the report should discuss this (maybe in an Annex with a special interview protocol for children).

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the ethical safeguards for children is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

Section D: Findings

Weighting = 20%

- Question 10. Do the findings clearly address all evaluation objectives and scope?
- Question 11. Are evaluation findings derived from the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence?
- Question 12. Does the evaluation assess and use the intervention's Results Based Management elements?



This element is particularly important due to the magnitude of relevant issues it unpacks.

Often, once the evidence gathering systems are found acceptable, findings, conclusions and lessons learnt are generally without significant flaws. However, in occasions, though data collection systems are designed well and explained, the actual evidence of field work and outputs in the field can also be relatively weak.

- Question 10. Do the findings clearly address all evaluation objectives and scope?
- Element 10.1. Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions and criteria
- Findings discuss several sources validating the same messages for each evaluation questions
- Evidence is used to address all of the evaluation's questions and criteria;
- Sufficient evidence was used (enough to prove the findings)
- The evidence is used systematically (using the evidence methodically and thoroughly to develop and show how the finding is true)
- All of the evaluation's questions and criteria were treated in this way.
- Summaries of key findings per question is an asset

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about findings' evidence is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

- Element I0.2. Reference to the intervention's results framework in the formulation of the findings
- Findings discuss several sources validating the Intervention's monitoring framework is discussed
- When possible, evaluation findings build on monitoring data

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about results' framework is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Question 11. Are evaluation findings derived from the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence.
- Element 11.1. The evaluation clearly presents multiple lines (including multiple time series) and levels (output, outcome, and appropriate disaggregation) of credible evidence.
- Intervention performance at output and outcomes levels
- How the data was analyzed and became evidence
- Credibility or potential gaps are discussed.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the lines of evidence is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 11.2. Findings are clearly supported by and respond to the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison.
- Clear disaggregation of positive and negative findings are presented
- Comparison between expected targets of the indicators or benchmarks and real-performance results should be presented and analyzed.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about judgement on findings is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 11.3. Unexpected effects (positive and negative) are identified and analyzed.
 - Unexpected effects are clearly described
 - Their causes and their effects are analyzed
 - Negative unexpected effects are present too
 - Or at least evaluators proof how they consistently looked for them in case they didn't find any.



No discussion of this issue in the report should be rated as “No”.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the unexpected results is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 11.4. The causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyze the logical chain (progression -or not- from implementation to results).
- Factors that have potentially caused the findings about the intervention’s performance.
- Comparison between these factors and the ones in the original theory of change elaborated during the design or rebuilt during the evaluation
- The whole logic of the program is reviewed.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the causal factors is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Question 12. Does the evaluation assess and use the intervention's Results Based Management elements?
 - Element 12.1. Clear and comprehensive assessment of the intervention's monitoring system (including completeness and appropriateness of results/performance framework -including vertical and horizontal logic; M&E tools and their usage)
 - Status and completeness of the monitoring system in place
 - Analysis of the monitoring system regarding its comprehensiveness, frequency, use of their outputs
 - Discussion of its overall value for monitoring outcomes performance by monitoring activities and outputs.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the M&E tools is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 12.2. (*) *In cases where a monitoring system exists and it is reliable for decision-making: clear and complete assessment of the use of monitoring data in decision making*
 - How the implementation team made advantage of the monitoring data
 - How it influenced some decisions regarding the intervention

Tip!

In case the report specifically mentions the lack of monitoring data, this element should be “Not rated”. If the report does not discuss it, the rating should be “No”.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the use of monitoring data is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Section E: Conclusions and Lessons Learned

Weighting = 15 %

- Question 13. Do the conclusions present an objective overall assessment of the intervention?
- Question 14. Are lessons learned correctly identified?



Tip!

This section should both summarize the findings; and state the consequences and implications.

There should not be a repetition of the main findings, but an overall analysis.

- Question 13. Do the conclusions present an objective overall assessment of the intervention?
 - Element 13.1. Clear and complete description of the strengths and weaknesses of the intervention that adds insight and analysis beyond the findings
- Overall description of the positive and negative findings
- Taking the findings one step beyond in the analysis.
- Done for each of the evaluation criteria and questions

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about insights from the findings is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 13.2. Description of the foreseeable implications of the findings for the future of the intervention (if formative evaluation or if the implementation is expected to continue or have additional phase)
 - Implications for the future of the intervention
 - Covering all evaluation criteria.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the implications of the findings is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

EXAMPLE

Some reports facilitate the identification of the origin of each conclusion by numbering every paragraph in the document and including a reference to the paragraphs where each conclusion is based.

Any other types of coding or mere verbal references can also be valid.

- Element I3.3. The conclusions are derived appropriately from findings
- Some kind of clear connected logic provides clear traceability about the main conclusions and the findings supporting them
- This is given for each main conclusion

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about traceability is provided
Not Rated	For some reason, this element is not applicable (Reason explained).



Lessons learned are required by UNICEF Standards.

However, if no Lessons Learned are present and the Conclusions in the report are satisfactory, the section can be rated as “Fair”.

- Question 14. Are lessons learned correctly identified?
 - Element 14.1. Correctly identified lessons that stem logically from the findings, presents an analysis of how they can be applied to different contexts and/or different sectors, and takes into account evidential limitations such as generalizing from single point observations.
- Clear generalization of some of the conclusions to a Lessons Learned level
- Identified Lessons Learned can be applied in different contexts

Section F: Recommendations

Weighting = 15%

- Question 15. Are recommendations well-grounded in the evaluation?
- Question 16. Are recommendations clearly presented?

- Question 15. Are recommendations well-grounded in the evaluation?
- Element 15.1. Recommendations are logically derived from the findings and/or conclusions
- The recommendations clearly derive from the conclusions.
- In occasions, some reports may introduce some kind of logical, traceable connection (codes, references) that make it easy to identify where recommendations come from, and they specify findings and conclusions they are based on.

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about recommendations' base is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

- Element 15.2. Recommendations are useful to primary intended users and uses (relevant to the intervention and provide realistic description of how they can be made operational in the context of the evaluation)
- Clear reference to the evaluation questions, objectives or purpose
- Primary and secondary users can find them useful
- Operational context described does not imply that recommendations made are not realistic

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about use of recommendations is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 15.3. Clear description of the process for developing recommendations, including a relevant explanation if the level of participation of stakeholders at this stage is not in proportion with the level of participation in the intervention and/or in the conduct of the evaluation
- Process used for developing the recommendations described
- Clear level of participation from some of the most relevant stakeholders groups included
- Relevant groups participated

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the process of elaborating the recommendations is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Finding	Recommendation	Priority	Action by	Comments
<p>Accountability to Affected People</p> <p>There was no systematic leadership or attention given to AAP issues in the response. Few actors were familiar with the term AAP and even fewer knew of the IASC's commitments. The potential for greater roll out was undermined by the lack of experienced practitioners or advocates.</p> <p>Involvement of affected people in all stages of the response project cycle was minimal and their expressed priorities (e.g. for education) often overlooked.</p> <p>A 2014 CwC working group had petered out by 2015. There were some excellent examples of good practice by individual agencies that can be models for response-wide improvements in AAP.</p>	AAP needs to be embraced and promoted by the HC and the HCT as something that can add great value to the response, in line with IASC guidance.	Critical	HC, HCT	<p>The need to prepare and operationalize an AAP action plan was already highlighted in the OPR recommendations. More work is needed in this area, however, and the recommendations here aim to guide the response to do so, building on what was suggested by the OPR.</p>
	Promote good practice in AAP and CwC through a formal network of AAP focal points or champions sitting within agencies, in clusters and in key hubs, and able to document, share and promote good practice. Scale up and expand upon existing models of good practice in CwC, such as the work of Internews.	Critical	OCHA, ICWG	
	Agencies and aid workers at all levels need to move beyond information collection and sporadic consultation with affected people and embrace the full range of AAP commitments, focusing in particular on the participation of affected people in the programme cycle.	Critical	All agencies	
	Go beyond community leaders and committees and engage in dialogue with ordinary affected people.	Important	All	
	Dedicate mobile support to states to build capacity and integrate AAP into all clusters through training and mentoring. Use secondments from established agencies, such as Internews, into clusters	Important	OCHA	
Dedicate staff and budget lines for AAP. Promote good practice in AAP and CwC through a formal network of AAP focal points or champions sitting within agencies, in clusters and in key hubs, that are able to document, share and promote good practice. Scale up and expand upon existing models of good practice in CwC, such as the work of Internews.	Important	OCHA, agencies, donors		

From report: Inter-Agency Humanitarian Evaluation of the Response to Crisis in South Sudan (HQ 2016/004)

- Question 16. Are recommendations clearly presented?
- Element 16.1. Clear identification of target group for action for each recommendation (or clearly clustered group of recommendations)
- Each recommendation specifies to which target group it is addressed

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about recommendations' target groups is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element I 6.2. Clear prioritization and/or classification of recommendations to support use
- All the recommendations specify their relative priority among each other
- They are classified somehow

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about recommendations' priorities is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Section G: Structure and Presentation

Weighting = 5%

- Question 17. Does the evaluation report include all relevant information?
- Question 18. Is the report logically structured?

- Question 17. Does the evaluation report include all relevant information?
 - Element 17.1. Opening pages include: Name of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents -including, as relevant, tables, graphs, figures, annexes-; list of acronyms/abbreviations, page numbers
- The opening pages of the report include:
 - Name of evaluated object
 - Timeframe of the evaluation
 - date of report
 - Location of evaluated object,
 - names and/or organization(s) of the evaluator(s),
 - name of organization commissioning the evaluation
 - table of contents -including, as relevant, tables, graphs, figures, annexes-;
 - List of acronyms/abbreviations, page numbers

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about these elements is provided in the first pages
Not Rated	For some reason, this element is not applicable (Reason explained).

Example

From report: Final evaluation report:
“Developing Community based
Services for Children with
Disabilities and their Families” (April
2010 –December 2013).
(Serbia 2013/010)

Name of the Project: Developing Community-based Services for Children with Disabilities and Their Families

Timeframe of the evaluation: July-November 2013

Date of the report: December 2013

Country of project: Serbia

Evaluators: Camelia Gheorghe and Ozren Runic

Name and address of the company the report has been commissioned by: Promeso Consulting srl, 14, Decebal Blvd., 030967 Bucharest, Romania

Name of the organisation commissioning the evaluation: UNICEF Serbia

Name of UNICEF staff contact point for the evaluation: Vesna Dejanovic, Project Officer Child Protection

EXAMPLE

Annexes

- Annex 1: Terms of Reference
- Annex 2: Evaluation Questions for each of the Evaluation Criteria
- Annex 3: Evaluation Matrix
- Annex 4: Details on Evaluation Methodology
- Annex 5: List of Persons Consulted at national level
- Annex 6: Overview of Types of Measures included
- Annex 7: Monitoring Framework of the NCYDP
- Annex 8: Context in Administrative Areas visited by the team
- Annex 9: Case studies of selected Initiatives
- Annex 10: Impact related Tables
- Annex 11: Province Specific Data on selected Performance Indicators
- Annex 12: References
- Annex 13: CV of the evaluators

- Element 17.2. Annexes should include, when not present in the body of the report: Terms of Reference, Evaluation matrix, list of interviewees, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology, copy of the results chain, information about the evaluator(s)".
- All these annexes are included in the report:
 - Terms of Reference,
 - Evaluation matrix,
 - list of interviewees,
 - list of site visits,
 - data collection instruments (such as survey or interview questionnaires)
 - list of documentary evidence
- Other appropriate annexes could include: additional details on methodology, copy of the results chain, information about the evaluator/s

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No annexes are provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Question 18. Is the report logically structured?
 - Element 18.1. The structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles)
 - The structure is easy to identify
 - The structure is easy to navigate
 - Numbered sections and clear titles and sub-titles

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No clear structure is followed.
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 18.2. Context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations
- Context, purpose and methodology are in the first part
- Findings are followed by conclusions, lessons learned and recommendations
- The sections follow a logical order

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No logical order is followed.
Not Rated	For some reason, this element is not applicable (Reason explained).

Section H: Evaluative Principles

Weighting = 15%

- Question 19. Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity?
- Question 20. Does the evaluation assess the extent to which the implementation of the intervention addressed gender, equity & child rights?
- Question 21. Does the evaluation meet UN SWAP evaluation performance indicators?

This question assesses the application of principles in evaluation itself.

- It's interpretation and application is intended to be guided only by UNICEF policies, not UN-wide policies (such as UN-SWAP).
- This is designed to allow comparison over time.

- Question 19. Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity?
 - Element 19.1. Reference and use of rights-based framework, and/or CRC, and/or CCC, and/or CEDAW and/or other rights related benchmarks in the design of the evaluation
 - Rights-based framework
 - CEDAW (Convention on the Elimination of All Forms of Discrimination against Women)
 - Convention on the Rights of Persons with Disability (CRPD)
 - DEVAW (Declaration on the Elimination of Violence against Women)
 - Regional protocols on gender based violence
 - Universal declaration of Human Rights

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the rights-based framework is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

- Element 19.2. Clear description of the level of participation of key stakeholders in the conduct of the evaluation, and description of the rationale for the chosen level of participation (for example, a reference group is established, stakeholders are involved as informants or in data gathering)
 - Stakeholders participate in design, implementation and monitoring
 - Stakeholders consulted/participated in deciding what to evaluate and how to evaluate it
 - Evaluation measures stakeholders’ group participation?
 - Evaluation assesses how participation benefits groups

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the stakeholders’ participation is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 19.3. Stylistic evidence of the inclusion of these considerations can include: using human-rights language; gender-sensitive and child-sensitive writing; disaggregating data by gender, age and disability groups; disaggregating data by socially excluded groups.
 - Use of human-rights language
 - Use of gender-sensitive and child-sensitive writing
 - Disaggregated data by gender and age groups
 - Disaggregated data by socially excluded groups
 - Disaggregated data by groups with special needs (displaced, pregnant, disabilities)

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No details about the gender sensitivity of the report is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

This question assesses whether the evaluation explores the application of principles in the intervention being evaluated.

- It's interpretation and application is intended to be guided only by UNICEF policies, not UN-wide policies (such as UN-SWAP).
- This is designed to allow comparison over time.

- Question 20. Does the evaluation assess the extent to which the implementation of the intervention addressed gender, equity & child rights?
 - Element 20.1. Identification and assessment of the presence or absence of equity considerations in the design and implementation of the intervention
 - Intervention discusses balance power relations between groups
 - Intervention discusses balance power relations within groups
 - Supports the empowerment of disadvantaged groups

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the equity is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

June 2017

- **Element 20.2. Identification and assessment of the presence or absence of gender analysis in the design and implementation of the intervention**
 - Explicit analysis of gender differences between men and women (and boys and girls)
 - Physical and physiological differences and how they affect each
 - Different roles and cultural norms and how they affect them
 - Risks and vulnerabilities due to those differences
 - Access to services due to those differences
 - Results and consequences due to those differences

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No gender analysis is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 20.3. Explicit analysis of the involvement in the object of rights holders, duty bearers, and socially marginalized groups, and the differential benefits received by different groups of children
 - Analysis of the rights holders involvement
 - Analysis of the duty bearers involvement
 - Analysis of the marginalized groups involvement
 - Benefits for each of the groups

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the involvement is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 20.4. Clear proportionality between the level of participation in the intervention and in the evaluation, or clear explanation of deviation from this principle (this may be related to specifications of the TORs, inaccessibility of stakeholders at the time of the evaluation, budgetary constraints, etc.)
- Analysis of relative importance of each group
- Analysis of their importance vs participation in the intervention
- Analysis of their importance vs participation in the evaluation

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No information about the participation is provided.
Not Rated	For some reason, this element is not applicable (Reason explained).

UNEG Guidance for Scoring the UN-SWAP EPI

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

It is important to note that no decimals should be provided in the scoring of criteria, only whole numbers.

Tip!

June 2017

- Question 21. Does the evaluation meet UN SWAP evaluation performance indicators?
 - The United Nations *System-wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP)* constitutes the first accountability framework for gender mainstreaming in the UN system. The UN-SWAP is composed of 15 performance indicators for tracking 6 main elements on gender mainstreaming: accountability, results based management, oversight, human and financial resources, capacity, and knowledge exchange and networking.
 - Specific guidance on implementing the UN-SWAP evaluation performance indicator (EPI) is available from UNEG.

Not Applicable	Missing	Approaches requirements	Meets requirements	Exceeds requirements
5a. Performance indicator is not relevant to a UN entity	5b. None of the UNEG gender-related norms and standards are met	5c. Meets some of the UNEG gender-related norms and standards	5d. Meets the UNEG gender-related norms and standards	5ei. Meets the UNEG gender-related norms and standards and 5eii. Demonstrates effective use of the UNEG guidance on evaluating from a human rights and gender equality perspective

<p>1. Gender Equality and the Empowerment of Women (GEEW) is integrated in the Evaluation Scope of analysis and Indicators are designed in a way that ensures GEEW-related data will be collected</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Analysis on how GEEW objectives and GEEW mainstreaming principles were included in the intervention design <input type="checkbox"/> How GEEW results have been achieved. Assessment of the extent to which an intervention has been guided by organizational and system-wide objectives on GEEW. <input type="checkbox"/> Indicators for the evaluation of the intervention should include GEEW dimensions and/or additional indicators are identified specifically addressing GEEW; <input type="checkbox"/> Mixed indicators (including quantitative and qualitative indicators) are preferred. 	<p>0=Not at all integrated. 1=Partially integrated. 2=Satisfactorily integrated. 3=Fully integrated.</p>
---	---	--

<p>2. Evaluation Criteria and Evaluation Questions specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.</p>	<ul style="list-style-type: none"> <input type="checkbox"/> GEEW dimensions are integrated into all Evaluation Criteria and questions <input type="checkbox"/> Criteria derived directly from GEEW principles are used (e.g. equality, participation, social transformation, inclusiveness, empowerment, etc.). 	<p>0=Not at all integrated. 1=Slightly integrated. 2=Satisfactory integrated. 3=Fully integrated.</p>
--	---	---

<p>3.A gender-responsive Evaluation Methodology, Methods and tools, and Data Analysis Techniques are selected.</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Triangulation of data is done to ensure that the voices of both women, men, boys and girls are heard and used <input type="checkbox"/> Additional time or resources (time, staff, funds) to implement a gender-responsive approach is considered and planned for, etc. <input type="checkbox"/> Mixed-method approach are preferred to make visible diverse perspectives and promotes <input type="checkbox"/> Data collection methods including, desk reviews, focus groups, interviews, surveys, etc. are identified and accompanying tools, e.g. questionnaires, observational tools, interview guides etc. developed integrating GEEW considerations (e.g. interview guides ensure that women and men are interviewed in ways that avoid gender biases or the reinforcement of gender discrimination and unequal power relations, etc.). <input type="checkbox"/> During data screening and data analysis, special attention is paid to data and information that specifically refer to GEEW 	<p>0=Not at all integrated. 1=Slightly integrated. 2=Satisfactory integrated. 3=Fully integrated.</p>
<p>4.The evaluation Findings, Conclusions and Recommendation reflect a gender analysis</p>	<p>The evaluation report’s findings, conclusion and recommendations should reflect a gender analysis.</p> <ul style="list-style-type: none"> <input type="checkbox"/> The evaluation report should also provide lessons/challenges/recommendations for conducting gender-responsive evaluation based on the experience of that particular evaluation 	<p>0=Not at all integrated. 1=Slightly integrated. 2=Satisfactory integrated. 3=Fully integrated.</p>

Section I: Executive Summary

Weighting = 5%

- Question 22. Can the executive summary inform decision-making?

- Question 22. Can the executive summary inform decision-making?
- Element 22.1. An executive summary is provided that is of relevant conciseness and depth for primary intended users
- A clearly separated Executive Summary is present in the report
- It addresses the evaluation primary users' needs of information

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No executive summary is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 22.2. Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key findings, key conclusions, key recommendations)
- Executive summary Includes:
 - Overview of the intervention
 - Evaluation purpose
 - Objectives
 - Intended audience
 - Evaluation methodology
 - Most important findings
 - Key conclusions
 - Key recommendations

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No executive summary is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

- Element 22.3. Includes all the necessary information to understand the intervention and the evaluation **AND** does not contain information not already included in the rest of the report
- Includes all the necessary information to understand the intervention and the evaluation
- It does not contain information not already included in the rest of the report

Yes	The report or annexes specify very clearly all this information
Mostly	Most of these elements can be found, but not all of them.
Partly	Only some of them can be found.
No	No executive summary is provided
Not Rated	For some reason, this element is not applicable (Reason explained).

Links to the key GEROS tools and further resources

CHAPTER 4: ASSESSMENT TOOLS

GEROS Evaluation Quality Assurance template

Click the icon to open the template,
or [download here](#)



Microsoft Excel Worksheet

June 2017

Evaluation Norms and Standards

Norms and Standards for Evaluation (2016)

<http://www.unevaluation.org/document/detail/1914>

UNEG Quality Checklist for Evaluation Reports

<http://www.unevaluation.org/document/detail/607>

Integrating Human Rights and Gender Equality in Evaluations

<http://www.unevaluation.org/document/detail/1616>

UNICEF Report Standards Checklist

https://www.unicef.org/evaluation/files/UNICEF_adapted_reporting_standards_updated_June_2017.pdf

UNICEF TOR Standards Checklist

https://www.unicef.org/evaluation/files/UNICEF_UNEG_TOR_Checklist_updated_June_2017.pdf

June 2017

UNEG Norms 2005	UNEG Norms 2016
Norm 1: Definition	Norm 1: Internationally agreed principles, goals and targets
Norm 2: Responsibility for evaluation	Norm 13: Responsibility for the evaluation function
Norm 3: Policy	Norm 12: Evaluation policy
Norm 4: Intentionality	Norm 2: Utility
Norm 5: Impartiality	Norm 5: Impartiality
Norm 6: Independence	Norm 4: Independence
Norm 7: Evaluability	-
Norm 8: Quality of evaluation	Norm 3: Credibility
Norm 9: Competencies for evaluation	Norm 10: Professionalism
Norm 10: Transparency and consultation	Norm 7: Transparency
Norm 11: Evaluation ethics	Norm 6: Ethics
Norm 12: Follow-up to evaluation	Norm 14: Evaluation use and follow-up
Norm 13: Contribution to knowledge building	-
	Norm 8: Human rights and gender equality
	Norm 9: National evaluation capacities
	Norm 11: Enabling environment

CONTACTS

For further help and support
evalhelp@unicef.org

UNICEF GEROS Manager
Ada Ocampo

Handbook prepared by ImpactReady
Joseph Barnes | Sara Vaca



A summary version of this handbook is also available from the Evaluation Office