



Global Evaluation Report Oversight System (GEROS)

Guidance Handbook



GEROS Handbook

Version 4

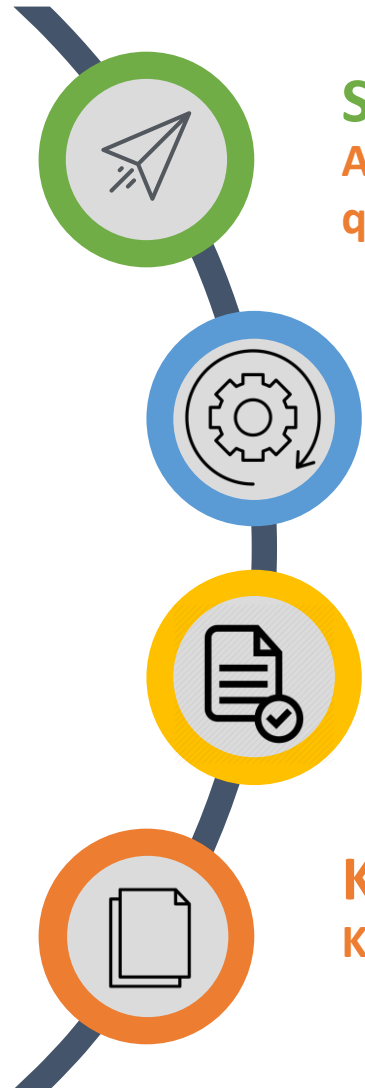
This handbook is designed to orient UNICEF staff, partners, evaluators and independent assessors to the Geros assessment tool and to the Geros system.

It is organized into three sections. The first provides an introduction to the Geros System. The subsequent sections describe the evaluation quality assessment process, and the standards that are used to review and rate each evaluation report. A list of key resources is then provided for further reference.

More information about Geros and the assessment tool is available from the Evaluation Office.



TABLE OF CONTENTS



SECTION 1: INTRODUCTION TO GEROS An overview of GEROS and the evaluation quality assessment process
SECTION 2: THE ASSESSMENT PROCESS How the assessment system works and what can be expected
SECTION 3: ASSESSMENT STANDARDS Overview of each element of the evaluation quality assessment
KEY REFERENCE DOCUMENTS Key reference documents

Acronyms

CCC	Core Commitments for Children	HRBAP	Human Rights Based Approach to Programming
CEDAW	Convention on the Elimination of all forms of Discrimination Against Women	LACR	Latin America and the Caribbean Region
CO	Country Office	MENA	Middle East and North Africa [Region]
CRC	Convention on the Rights of the Child	PPP	Programme Planning Process
EAPR	East Asia and Pacific Region	OECD-DAC	Organisation for Economic Cooperation and Development – Development Assistance Committee
ECAR	Europe and Central Asia Region	RBM	Results Based Management
EISI	Evidence Information Systems Integration (Database)	RO	Regional Office
EPI	Evaluation Performance Indicator (for UN-SWAP)	SAR	South Asia Region
EQA	Evaluation Quality Assessment	SDG	Sustainable Development Goal
ESAR	Eastern and Southern Africa Region	ToC	Theory of Change
GEEW	Gender Equality and Empowerment of Women	UNEG	United Nations Evaluation Group
GEROS	Global Evaluation Report Oversight System	UN SWAP	United Nations System-wide Action Plan
HQ	Headquarters [New York]	WCAR	West and Central Africa Region



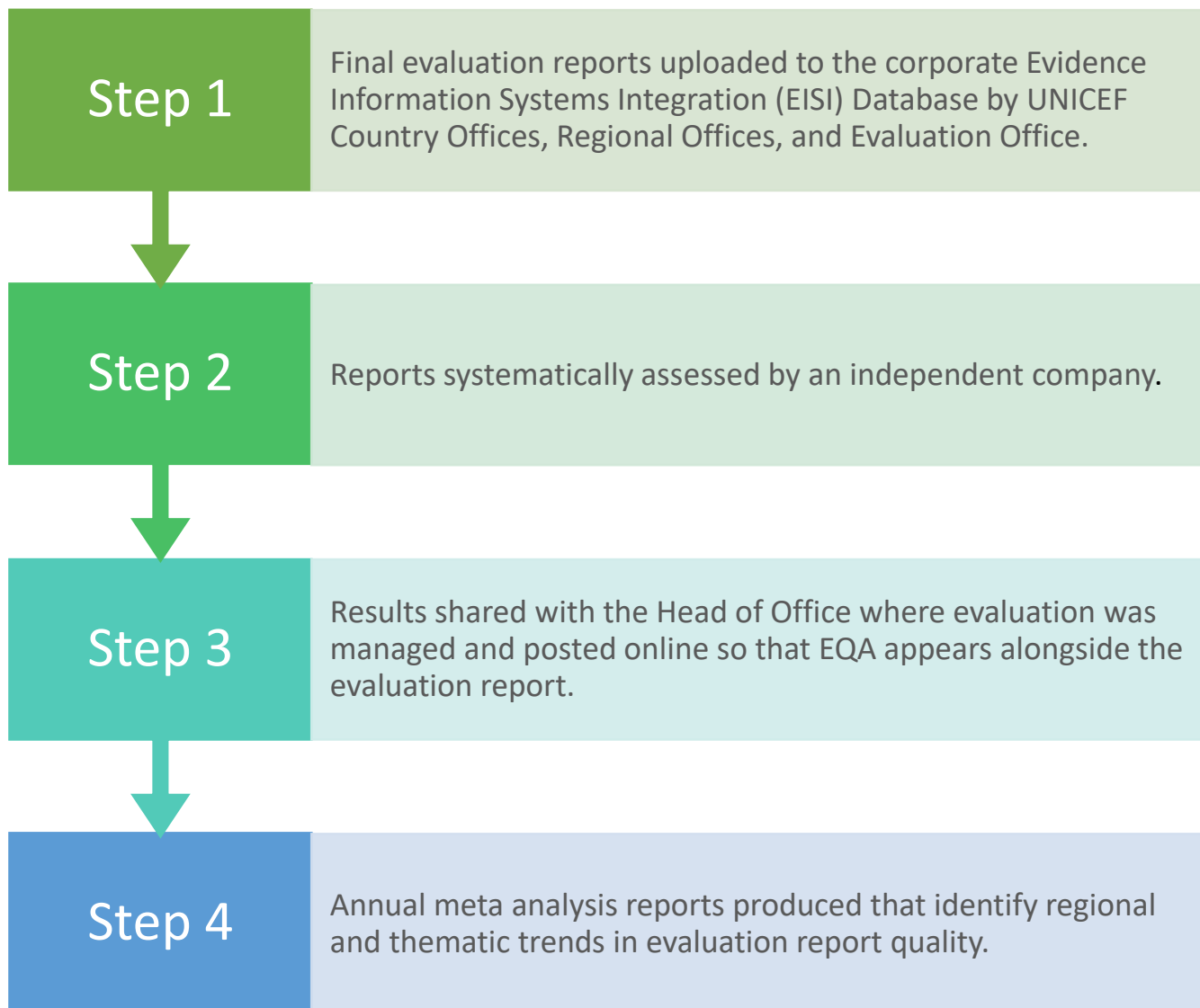
Section 1: Introduction to GEROS

An overview of GEROS and the evaluation quality assessment process

What is GEROS?

The Global Evaluation Reports Oversight System (GEROS) is an organization-wide system that aims to support strengthening of the evaluation function to meet and exceed United Nations Evaluation Group (UNEG) norms and standards, UN System Wide Action Plan on gender equality (UN-SWAP), and other UNICEF commitments (including equity, human-rights based and disability inclusive approaches).

UNICEF's evaluation quality assessment (EQA) system



GEROS Aims

A management tool that:

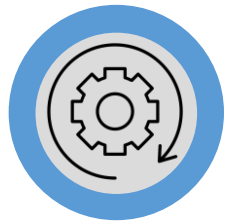
- Helps ensure accountability and promote use of robust evaluation evidence
- Provides an incentive for evaluation managers and evaluators to follow good practice
- Gives constructive feedback to help improve credibility and use of future evaluations

GEROS Objectives

1. Create an 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
2. Provide feedback that leads to stronger evaluation capacity of UNICEF and partners
3. Enable UNICEF and partners to be more knowledgeable about what works, where and for who.

GEROS Quality Rating Scale

Overall Rating		Implication
•••••	Exceptional	Exemplifies UNICEF/UNEG standards for evaluation reports. Decision makers may use the evaluation with a high degree of confidence
••••-	Highly Satisfactory	Exceeds UNICEF/UNEG standards for evaluation reports. Decision makers may use the evaluation with a high degree of confidence
•••--	Satisfactory	Meets UNICEF/UNEG standards for evaluation reports. Decision makers may use the evaluation with confidence
••---	Fair	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
•----	Unsatisfactory	Does not sufficiently meet the UNICEF/UNEG standards for evaluation reports. Decision makers cannot rely on the evaluation.
	Not Rated	Refers to aspect of the evaluation that was not rated for a legitimate reason. Normally this does not affect the quality of evaluation report.



Section 2: The Assessment Process

How the assessment system works and what can be expected

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 reviewed using the UNICEF-Adapted UNEG Evaluation Report Standards to assess the following core elements:

- Background of the evaluation is clearly and fully described
- Evaluation's purpose, objectives and scope are fully explained
- Methodology is appropriate, sound and in line with the ethical standards and procedures
- Findings, conclusions and recommendations are based on evidence, sound analysis and are useful
- Lessons learned support organizational learning
- Report is well structured, logical and clear
- UN-SWAP evaluation performance indicator for integration of gender equality is met or exceeded

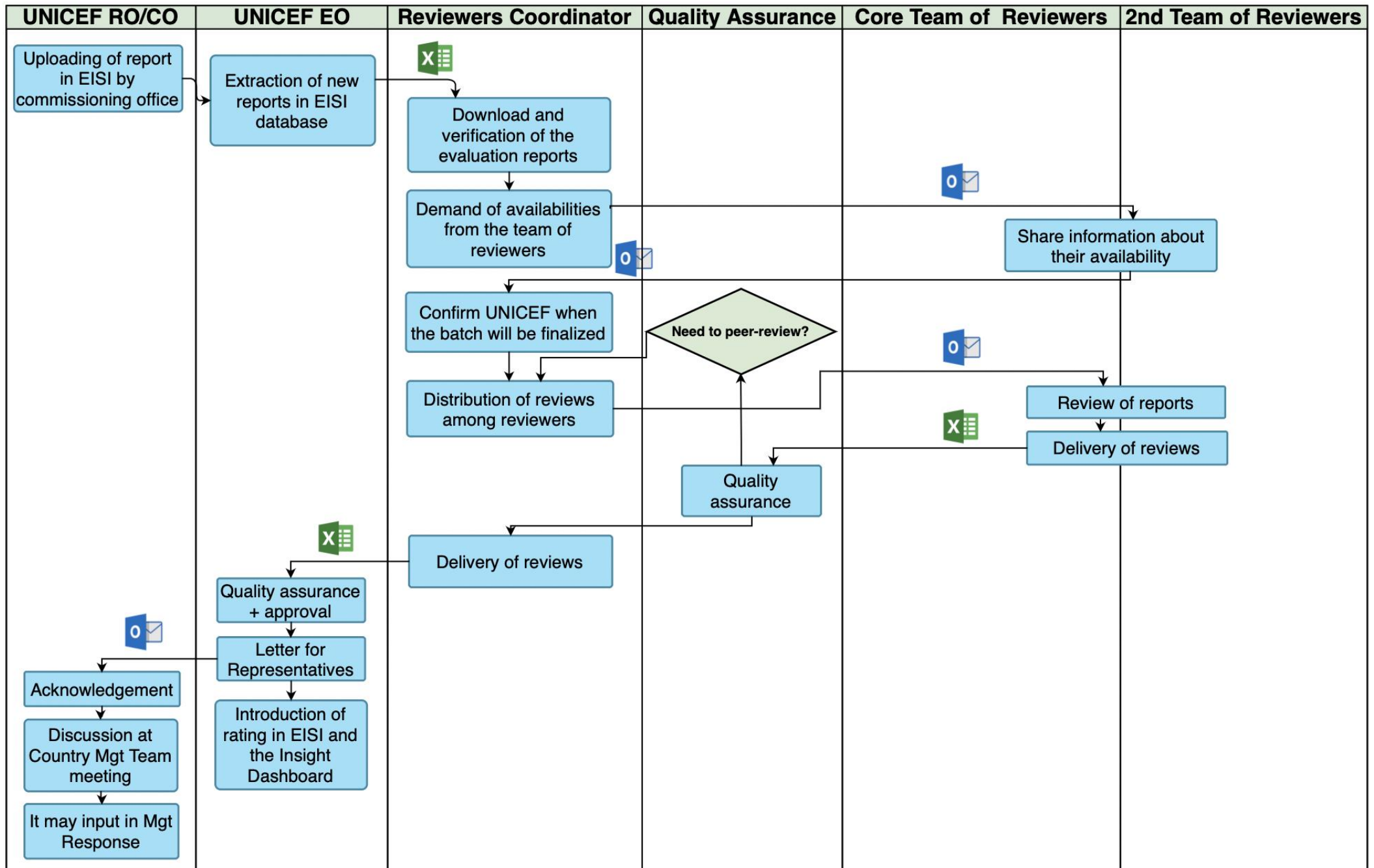
Standards for a good quality GEROS assessment

A useful evaluation quality assessment (EQA) supports organizational learning and building of evaluation capacity.

A good quality EQA is:

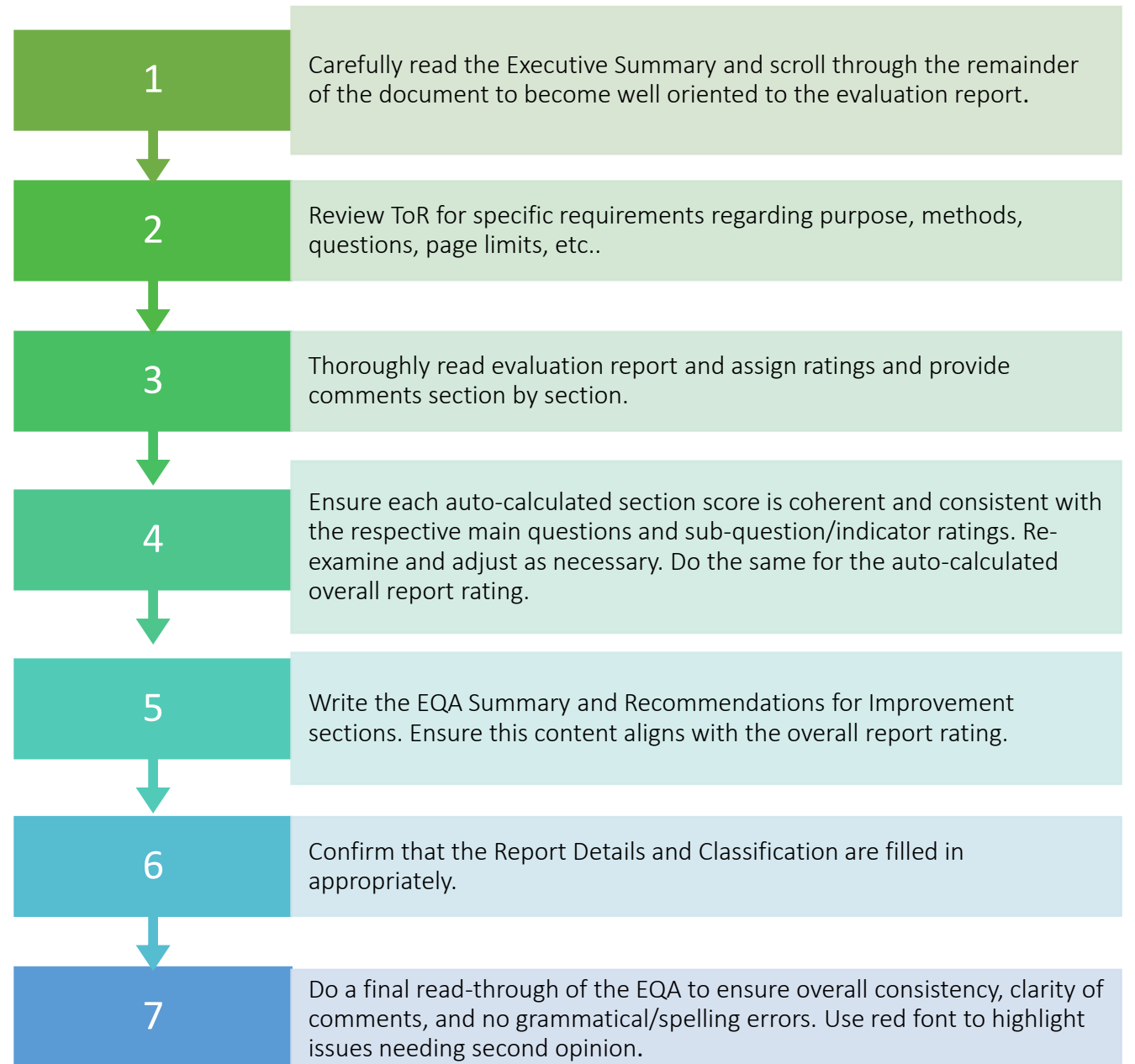
- Timely, submitted within the agreed timeframe
- Complete and in the original language of the evaluation report
- Accurate and reliable, giving examples where relevant. Does not second-guess the evaluation process itself.
- Consistent between quantitative and qualitative ratings, and across the evaluation 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

Assessment process workflow



Steps for conducting the EQA

UNICEF is particularly interested in the Evaluation Principles section. Reviewers are to ensure there is not just a passing reference made to key issues such as ethics, gender equality, disability inclusion, equity and human rights-based approach, but that they are actually integrated into the report. It is recommended that notes for this section be taken throughout the review of the report.



Recent changes to EQA template

The EQA template is periodically updated to respond to changes in UNEG guidelines and UNICEF's commitment to advancing good evaluation practice.

Changes are discussed and agreed at the Annual Review Workshops for GEROS prior to being implemented. On occasion, minor revisions may be made by the Evaluation Office primarily for the purpose of clarifying the intent of the EQA questions or indicators.

Changes implemented over the past two years include expectations that evaluations will:

1. Use rights-based language to reflect the UN's human rights-based approach to development: [HRBA Approach \(UNSDG\)](#). This entails referring to those the intervention is intended to assist as 'rights-holders' and the institutions obligated to fulfill the holder's rights as 'duty-bearers'.
2. Use the OECD-DAC Criteria most relevant to the evaluation object. [OECD-DAC Criteria \(2019\)](#) no longer require all evaluations to cover all criteria.
3. Align with the [UN Disability Inclusion Strategy \(2020\)](#). A disability inclusive evaluation considers disability in addition to the principle of Leave No One Behind.
4. Adhere to UNEG Ethical Principles [UNEG Ethical Guidelines for Evaluation \(2020\)](#). Good practice is for evaluation team members to acknowledge and sign the [UNEG Pledge of Ethical Conduct for Evaluators](#) in addition to explaining how the UNEG Code of Ethics for Evaluators has been upheld.

Going forward, UNICEF is also interested in seeing evidence of more robust and innovative evaluation methods.

The rating scale

Each indicator receives a rating and an explanatory comment to justify the rating.

The EQA template includes 22 questions, each having multiple indicators (sub-questions). Indicators are rated on a 4 point scale:

Yes	The indicator is fully met
Partially	Some aspects of the indicator have been met but not all
No	There is no or only minimal evidence of the indicator being met
Not Rated	The indicator is not relevant to the evaluation report

Providing constructive feedback

The review process is iterative. The reviewer will need to move back and forth between the sections to add further explanations and to check and revise ratings and comments where needed.

1. Comments on Ratings should briefly indicate how the report met or did not meet each criteria. It is helpful to cite illustrative examples.
2. EQA Summary box should provide a concise overview of how the evaluation performed according to the UNICEF and UNEG standards. It should highlight the topline strengths and weaknesses of each section, note the methodology used, highlight best practices and any factors that contributed to the complexity of the evaluation process.
3. Recommendations for Improvement should be specific to the sections of the report where shortcomings were identified. Where weaknesses are significant, resources should be cited that may be useful for future evaluations. As relevant, suggestions should also be made for increasing methodological rigour.

Style of comments

- Write as a peer, putting yourself in the shoes of the evaluator: aim to 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 not including the ToR as an annex).
- Be a guide to the UNEG standards: use phrases such as "the UNICEF/UNEG standards require that..." particularly in the EQA Summary and Areas for Improvement boxes 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.

Weighting of quality assessment ratings

The overall rating for an evaluation report is based on a weighted aggregation of the scores from the nine EQA sections.

Each section has a percentage score that is based on a set of guiding questions. Each question has several criteria that are rated on a 4 point scale. The criteria are equally weighted and are auto-calculated to provide the section score.

The justification for each criterion rating is included in written comments to support transparency.

Section	Weighting (%)
Executive Summary	5
Background	5
Purpose, Objectives and Scope	5
Design and Methodology	20
Findings	25
Conclusions and Lessons	10
Recommendations	15
Structure and Presentation	5
Evaluation Principles	10
TOTAL	100

The EQA template and weighting are discussed at the Annual Review Workshops for GEROS. These rating were agreed as of July 2020.

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 reviewing firm

Classification of evaluation report (UNICEF taxonomy)

- Management of evaluation (managerial control and oversight of evaluation decisions); UNICEF goal areas correspondence (alignment with strategic plan priorities); Evaluation object; Evaluation type; Evaluation strategy; Evaluation design; Evaluation level; Geographic scope; Primary SDG(s) covered

Main assessment

- Sections A-I; Section Rating; Criteria Ratings; Comments on Ratings

Overall performance

- Overall rating; Executive summary; Recommendations for Improvement

Data sheet (auto-generated)

- All data provided in single row for inserting in database for analysis

EQA reports and follow-up

EQA report: The EQA process generates 4 products for use by UNICEF staff

1. A full copy of each independent assessment (Excel and PDF)
2. A letter to the head of office responsible for each evaluation explaining the EQA result
3. The UNICEF Insight Evaluation Dashboard
4. An annual meta analysis report

Management Response: These 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.



Section 3: Assessment Standards

Overview of each element of the evaluation quality assessment

Meta-data (used for analysing overall trends)

Assessment details

- Title of evaluation report
- Report sequence number
- Region
- Year of report
- Office
- Coverage (countries)
- ToRs present
- Date of Review
- Name of review firm

Classification of the evaluation

- Management of Evaluation
- UNICEF Goal Area Correspondence
- Evaluation object
- Evaluation level
- Geographic Scope
- Primary SDGs covered
- 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

EXECUTIVE SUMMARY

Question 1.

Can the executive summary inform decision-making?

i

Is clearly presented, serves as a standalone document useful for informing decision making, and is of relevant conciseness and depth for key users (Maximum of 5 pages unless otherwise specified in ToR).

ii

Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key conclusions on findings, lessons learned if requested, key recommendations) as per the ToR.

iii

Includes all significant information needed to understand the intervention and the evaluation AND does not introduce new information from what is presented 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

BACKGROUND	
Question 2.	Is the object of the evaluation clearly described?
i	Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status.
ii	Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) 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, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).
Question 3.	Is the context of the intervention clearly described?
i	Clear and relevant description of the context of the intervention (i.e. relevant policy, socio-economic, political, cultural, power/privilege, institutional, international factors) and how context relates to the implementation of the intervention.
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.
iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.
Question 4.	Are key stakeholders, their relationships and contributions clearly identified?
i	Identification of implementing agency(ies), development partners, right holders, and additional duty bearers and other stakeholders; and of linkages between them (e.g., stakeholder map) (if relevant).
ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.

Assessment sections, questions and indicators

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

EVALUATION PURPOSE, OBJECTIVES AND SCOPE	
Question 5.	Is the purpose of the evaluation clearly described?
i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.
Question 6.	Are the objectives and scope of the evaluation clear and realistic?
i	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 (if applicable).
ii	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, if applicable, 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).
Question 7.	Is the theory of change, results chain or logic well articulated?
i	Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.
ii	Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).
iii	For theory-based evaluations, the theory of change or results framework is assessed, and if requested in the ToR, it is reformulated/improved by the evaluators.

Assessment sections, questions and indicators

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

EVALUATION DESIGN AND METHODOLOGY (continued)	
Question 9. (continued)	Does the report specify adequate methods for data collection, analysis, and sampling?
iv	Clear and complete description of the methods of analysis.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.
vi	Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).
Question 10.	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:
i	Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or UNEG Ethical Principles.
ii	Description of ethical safeguards for participants appropriate for the issues relevant to methodology and how they are applied (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm). 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

EVALUATION DESIGN AND METHODOLOGY (continued)	
Question 8.	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? <i>UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender, human rights based approach, and disability inclusion (these can be mainstreamed into other criteria). Humanitarian evaluations should also consider Coverage; Connectedness; Coordination; Protection; Security.</i>
i	Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.
ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.
Question 9.	Does the report specify adequate methods for data collection, analysis, and sampling?
i	Evaluation design and set of methods is relevant and adequately robust for the evaluation's purpose, objectives and scope; and are fully and clearly described.
ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

✓ SECTION E

SECTION F

SECTION G

SECTION H

SECTION I

EVALUATION FINDINGS	
Question 11.	Do the findings clearly address all evaluation objectives and scope?
i	Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions, sub-questions and criteria.
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.
Question 12.	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.
i	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident through the use of multiple data sources.
ii	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 as relevant for each question.
iii	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).

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

SECTION E

✓ SECTION F

SECTION G

SECTION H

SECTION I

EVALUATION CONCLUSIONS & LESSONS LEARNED

Question 14. Do the conclusions clearly present an objective overall assessment of the intervention?

i

Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).

ii

Conclusions are derived appropriately from findings, and present a picture of the strengths and limitations of the intervention that adds insight and analysis beyond the findings.

Question 15. Are logical and informative lessons learned identified? [N/A if lessons are not presented and not requested in ToR]

i

Identified lessons stem logically from the findings, have wider applicability and relevance beyond the object of the evaluation.

ii

Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

SECTION E

SECTION F

✓ SECTION G

SECTION H

SECTION I

RECOMMENDATIONS

Question 16. Are recommendations well grounded in the evaluation?

i

Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.

ii

Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.

iii

Process for developing the recommendations is described, and includes the Involvement of duty-bearers, as well as rights holders when feasible (or explanation given for why they were not involved).

Question 17. Are recommendations clearly presented?

i

Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

SECTION E

SECTION F

SECTION G

✓ SECTION H

SECTION I

REPORT STRUCTURE AND PRESENTATION

Question 18. Does the evaluation report include all relevant information?

i

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.

ii

Annexes include, if not in report body: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework, 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, results chain, information about the evaluator(s).

Question 19. Is the report logically structured?

i

Structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles, well formatted).

ii

Structure accords to UNICEF guidelines for evaluation reports: context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.

Assessment sections, questions and indicators

SECTION A

SECTION B

SECTION C

SECTION D

SECTION E

SECTION F

SECTION G

SECTION H

✓ SECTION I

EVALUATION PRINCIPLES	
Question 21.	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?
i	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.
ii	Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).
iii	Language is empowering and inclusive, avoiding gender, heterosexual, age, cultural and religious bias, among others; use terminology of rights holders and duty bearers; data is disaggregated by marginalized group; differential results are assessed (distribution of results across different groups).
iv	Evaluation assesses the extent to which the implementation of the intervention addresses child rights and Leave No-one Behind (gender and other excluded and marginalized groups). It is disability inclusive.
Question 22.	Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards)
i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.

SWAP rating guidance

i GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.

- a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?
 - b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?
 - c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?
 - d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?
-

SWAP rating guidance

ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.

- a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?
 - b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations?
 - c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?
 - d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?
 - e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?
-

SWAP rating guidance

iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.

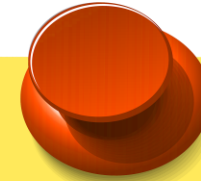
- a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?
 - b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?
 - c. Are unanticipated effects of the intervention on human rights and gender equality described?
 - d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area?
-

How UN-SWAP scores are determined

All evaluations are scored according to three evaluation performance indicators (EPI) set out in the United Nations System-wide Action Plan on Gender. The EPIs are included under EQA question 22. The scores reflect the extent that the evaluation report meets gender-related UNEG Norms & Standards for integrating human right and gender equality (GEEW). The allocation of points is explained in the adjacent note.

The next slide provides detailed guidance on the elements that raters should take into account in making their assessment.

Further information can be found on the [UNWomen](#) website.



Scoring system for the UNSWAP EPI

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

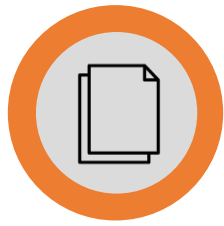
1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

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.



Key reference documents

ALNAP Evaluation of Humanitarian Action Guide (2016)	https://www.alnap.org/help-library/evaluation-of-humanitarian-action-guide
OECD/DAC Evaluation Criteria (2019)	https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm
UNEG Ethical Guidelines for Evaluation (2020)	http://www.unevaluation.org/document/detail/2866
UNEG guidance on Integrating Human Rights and Gender Equality in Evaluations (2014)	http://www.unevaluation.org/document/detail/1616
UNEG Norms and Standards for Evaluation (2016)	http://www.unevaluation.org/document/detail/1914
UNICEF Adapted Reporting Standards (2017)	https://www.unicef.org/evaluation/files/UNICEF_adapted_reporting_standards_updated_June_2017.pdf
UN Disability Inclusion Strategy (2020)	https://www.un.org/en/content/disabilitystrategy/
UNSDG Human Rights-Based Approach	https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach

