

GUIDANCE NOTE

GLOBAL EVALUATION REPORT ASSESSMENT AND ANALYSIS SYSTEM (GERAAS)



GUIDANCE NOTE

GLOBAL EVALUATION REPORT
ASSESSMENT AND ANALYSIS
SYSTEM (GERAAS)



Independent Evaluation Service (IES)
Independent Evaluation and Audit Services (IEAS)
UN Women

New York, August 2019



TABLE OF CONTENTS

I. INTRODUCTION	1
II. WHAT IS THE UN WOMEN INDEPENDENT EVALUATION REPORT QUALITY ASSESSMENT?	1
III. USERS OF THE GUIDANCE	2
IV. APPLICABILITY	2
V. APPROACHES AND METHODS	2

VI. STEPS FOR REVIEWING AND RATING INDIVIDUAL EVALUATION REPORTS	2
Stage 1: Report classification and filtering	3
Stage 2: Report review and rating	3
Stage 3: Rating the parameters and overall score of the report	4
Stage 4: Sharing final reviews and executive feedback on individual reports	5

VII. META-EVALUATION REPORT	5
VIII. ROLE OF IES	6

ANNEX 1: UN WOMEN GLOBAL EVALUATION QUALITY ASSESSMENT AND RATING MATRIX	7
---	----------

I. INTRODUCTION

The evaluation function is a priority for UN Women as it ensures institutional accountability, learning and informed decision-making. This is achieved, inter-alia, through high-quality evaluations that inform evidence-based decision-making to help advance achievement of gender equality and women's empowerment. The UN Women Evaluation Policy gives greater pertinence to the credibility and quality of evaluation processes and products.

While the UN Women Independent Evaluation Service (IES) provides leadership and quality assurance of the evaluation function, the planning, commissioning and management of most evaluations are decentralized. Approximately 90 per cent of annual evaluation reports are decentralized, commissioned by Business Units at the decentralized level and headquarters divisions.

To ensure good quality and credible evaluations, particularly at the decentralized level, in 2013 IES established the Global Evaluation Reports Assessment and Analysis System (GERAAS). Within the framework of the Global Evaluation Strategy 2018–2021, the GERAAS guidance and the Evaluation Quality Assessment (EQA) matrix have been revised to further enhance the quality and credibility of evaluations cognizant of UN Women's institutional maturity. The GERAAS guidance and the EQA are also aligned with the revised UNEG norms and standards (2016).

Through the GERAAS, all completed evaluations within UN Women are independently assessed, and the rating and review feedback are posted in UN Women's publicly accessible database, the [Global Accountability and Tracking of Evaluation Use \(GATE\)](#).

II. WHAT IS THE UN WOMEN INDEPENDENT EVALUATION REPORT QUALITY ASSESSMENT?

The GERAAS is an organization-wide system established to assess the quality of UN Women's evaluation reports. The GERAAS is a central tenet of the IES strategy to strengthen the quality, transparency, credibility and utility of UN Women evaluations. The GERAAS uses the [UNEG evaluation report standards](#) as a basis for review and assessment, while ensuring specific standards relevant to UN Women.

The GERAAS is complemented by a range of evaluation quality assurance mechanisms, which provide quality assurance of the evaluation process and products. In the long term, the GERAAS will constitute a mechanism for near time independent assessment and feedback of the quality of evaluation reports provided to offices to allow quality improvement of reports. Details of quality assurance standards aligned with different stages of the evaluation phases are provided in the UN Women Evaluation Handbook on "[How to manage gender-responsive evaluations](#)".

III. USERS OF THE GUIDANCE

The GERAAS guidance and EQA matrix seek to provide more clarity and guidance on the quality standards required for evaluation reports. The standards should be used by evaluation managers, evaluators and independent assessors to ensure that the evaluations produced by UN Women are high quality and credible. By providing constructive feedback to commissioning offices, IES aims to provide an incentive for evaluation managers and evaluators to improve the quality of future evaluation reports. In this sense, the EQA matrix also serves as a self-assessment tool and a means of communication between all actors involved in evaluation (evaluation managers, consultants/experts, evaluation reference group, Regional Evaluation Specialists etc.).

IV. APPLICABILITY

The GERAAS is an organization-wide system and the pre-defined standards should be applied and used for all types of evaluations (corporate, decentralized and joint evaluations) commissioned by UN Women headquarters Business Units, Decentralized Offices and IES.

V. APPROACHES AND METHODS

GERAAS uses the UNEG and United Nations System-Wide Action Plan for Gender Equality Evaluation Performance Indicator (UN-SWAP EPI) evaluation report standards as a basis for assessment and rating, while ensuring specific standards relevant to UN Women. The assessment acknowledges the wide variety of contexts in which evaluation reports are produced and the resources available. In doing so, the GERAAS seeks to focus on developing constructive insights and capacity building in offices to improve the quality and usefulness of future evaluations. Through its annual meta-evaluation report, the GERAAS also aims to contribute towards capitalizing on the knowledge produced from evaluation reports and on capturing trends with different parameters across regions.

The GERAAS assesses final evaluation reports and accompanying annexes posted in GATE. To ensure credibility and objectivity, the quality assessment of final evaluation reports is undertaken by an external and independent firm or an individual selected through an open bidding process. The rating needs to give greater weight to the quality of the parameters and subrubrics rather than the extent to which the parameters or subrubrics are present in the body of the report.

VI. STEPS FOR REVIEWING AND RATING INDIVIDUAL EVALUATION REPORTS

The assessment process includes reviewing reports against pre-defined standards, completing the EQA matrix (including UN-SWAP Scorecard) and the executive feedback for each report. The EQA matrix is composed of eight parameters:

1. Object and context of the evaluation
2. Evaluation purpose, objectives and scope
3. Evaluation methodology
4. Findings

5. Conclusions and lessons learned
6. Recommendations
7. Gender and human rights considerations (UN-SWAP EPI)
8. Report presentation.

In general, the independent assessment and rating process consist of four main stages:¹

Stage 1: Report classification and filtering

This comprises basic information such as title, region/country, type, costs, geographic and thematic coverage, stage/timing and management of the evaluation.

Stage 2: Report review and rating

Every assessment begins with a thorough reading of the evaluation report. In particular, the assessor is expected to:

- Review the report and accompanying annexes in detail using the GERAAS EQA matrix.
- Perform a quality review of each report based on the EQA matrix template. The assessment matrix provides eight pre-defined parameters including the UN-SWAP Evaluation Performance Indicator Scorecard.
- Conduct the review based on what is written in the evaluation report to ensure consistency in the rating of each parameter and the overall report.

General information on the EQA matrix

- Each parameter/section has been associated with a weighting (or a multiplying factor) which is proportionate to and illustrates its relative importance to the overall quality of the report.
- While all parameters are important, the assessment gives more weight to findings, conclusions, methodology and recommendations. Taken together, these four criteria constitute nearly 70 per cent of the total quality score.
- Each parameter is further disaggregated into 29 rubrics. Under each quality parameter, the relative importance of the subrubrics varies, and the score given to each subrubric is weighted.
- To support transparency in the rating, the EQA tool displays the criteria weight given to each rubric within the parameter, the weighted increments and the raw point score against all 29 rubrics.
- Usually, it is important for a report written in Spanish or French, for example, to be reviewed by a native speaker or someone with a high degree of proficiency.

¹ This section and the steps outlined are provided mainly to guide the external assessor to independently review and rate the quality of final evaluation reports.

[Table 1: The eight parameters/sections and associated weighting](#)

Parameter/section	Parameter Weight (%)
1: Object and context	5
2: Purpose, objectives and scope	5
3: Methodology	15
4: Findings	20
5: Conclusions and lessons learned	20
6: Recommendations	15
7: Gender equality and human rights (UN-SWAP)	10
8: Report presentation	10

Stage 3: Rating the parameters and overall report score

- Each parameter/section has several rubrics that are weighted against their relative importance within the parameter. In the matrix, each rubric is scored as *Fully (3)*, *Mostly (2)*, *Partially (1)* or *Not at all (0)*.
- Based on the rating of the subrubrics under each parameter, the overall rating for each parameter is automatically assigned by the aggregation of the weighted score as *Very good*, *Good*, *Fair* or *Unsatisfactory*.
- Provide executive feedback under each parameter. Comments should focus on the section overall rather than on a particular rubric. The justification for the overall rating of each section/parameter should be included in the written feedback space provided under each parameter (*more guidance on feedback to commissioning offices is provided under Stage 4 below*).
- The UN System-Wide Action Plan for Gender Equality (UN-SWAP Evaluation Performance Indicator) criteria are rated according to the methods set by UNEG, with results integrated into the GERAAS rating. The UN-SWAP Evaluation Scorecard is a reporting tool organized around three scoring criteria that capture the overall elements related to integrating gender equality dimensions in evaluation reports. For this specific subparameter (Section 7 of the matrix), each criteria is rated as *Fully integrated (3)*, *Satisfactorily integrated (2)*, *Partially integrated (1)* or *Not at all integrated (0)*. Based on the rating of the criteria, the overall rating for the parameter is automatically assigned by the aggregation of the score, as *Meeting requirements (7 and above)*, *Approaching requirements (4 and above)*, and *Missing requirements (less than 4)*. UNEG endorsed the [technical note and scorecard on the UN-SWAP EPI](#) which should be referenced for further details.
- Once all rubrics are assessed, the total weighted score and overall rating are automatically generated in *Part III The Overall Rating* of the matrix. The overall rating of an evaluation is based on an aggregation of the weighted values of the parameters and the subrubrics against a four-scale rating, which is *Very good (85% and above)*, *Good (65% and above)*, *Fair (50% and above)* or *Unsatisfactory (less than 50%)*.

- The overall rating and assessment of the strengths and weaknesses of the evaluation report gives an indication of the relative reliability of the results and determines the extent to which the report can be used with confidence to feed into future programming and to serve other purposes.

[Table 2: The quality rating scale²](#)

GERAAS Quality Rating Scale	UN SWAP Evaluation Performance Rating	Implication	GERAAS score	Description of UN-SWAP scores
Very good	Meets requirement	The report can be used with confidence and is considered a good example.	3 – Fully	3 – Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation.
Good	Approaches requirements	The report adheres to UN Women evaluation standards and can be used with confidence.	2 – Mostly	2 – Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but improvements could still be made.
Fair	Misses requirement	The report meets certain standards, but some elements are missing or inadequately addressed. The report should be used with caution and substantive improvements in some areas are needed.	1 – Partially	1 – Partially integrated. Applies when some minimal elements are met but further progress and remedial action are needed to meet the standard required.
Unsatisfactory		The report has serious limitations and therefore cannot be used with any level of confidence.	0 – Not at all	0 – Not at all integrated. Applies when none of the elements under a criterion are met.

Stage 4: Sharing final reviews and executive feedback on individual reports

Inherent in the GERAAS is provision of specific executive feedback to commissioning offices about the quality of evaluation reports they managed. This is mainly designed to strengthen internal evaluation capacity by providing practical recommendations to improve future evaluations and to inform assessment of external consultants’ performance who might be hired for future evaluations. Where possible, feedback should be provided in the original language of the report.

VII. META-EVALUATION REPORT

The GERAAS includes an annual analysis of the overall quality of a series of evaluation reports. This meta-evaluation report focuses on presenting insights from analyses that can most usefully support future action to improve the quality of evaluation reports. The report needs to illustrate and summarize the strengths and possible weaknesses of evaluations in the relevant period to develop recommendations for future evaluation work. The meta-evaluation report also identifies good practices.

² GERAAS uses a four-scale rating system based on the extent to which reports meet the assessment criteria.

VIII. ROLE OF IES

IES oversees, coordinates and supports the GERAAS review process from the selection of consulting firm/individual consultant to finalization and provision of executive feedback to the offices concerned. While the external reviewer primarily takes sole responsibility for ensuring the completeness, quality and adherence to established standards, IES closely monitors the consistent application of approaches, rating and finalization of the exercise in a timely manner.

IES presents the findings of the review at the Annual Session of the Executive Board and to senior managers and the Global Evaluation Advisory Committee. The report is also shared with the headquarters divisions, Regional Offices and Country Offices concerned to improve the quality and utility of evaluations by highlighting the strengths, good practices and areas that require improvement. The report is posted on the GATE along with the rating and executive feedback system, which allows access to the general public. This contributes to the transparency and credibility of UN Women when reporting on its performance.

ANNEX 1. UN WOMEN GLOBAL EVALUATION QUALITY ASSESSMENT AND RATING MATRIX (sample)

Independent Evaluation and Audit Services (IEAS)
UN WOMEN Global Evaluation Quality Assessment and Rating

Rating Scale	Very Good	Good	Fair	Unsatisfactory	Reviewer Guidance :
Rating explanation	The report can be used with high level of confidence and is considered a good example.	The report can be used with certain degree of confidence.	Partially meets requirements with some missing elements. The report can be used with caution.	Misses out the minimum quality standards.	<ul style="list-style-type: none"> - Overall reports are rated against a 4-point scale (Very Good, Good, Fair and Unsatisfactory), which is an aggregated rating of eight parameters. - Each overarching parameter is rated against a 4-point scale (Fully, Mostly, Partially and Not at all). - Parameters such as evaluation methodology, findings, conclusions and recommendations are given more weight. - Executive feedback - provide summary of the extent to which the report meets or fails to meet the criteria provided under each parameter. Please also include suggestion on how to improve future evaluation practice. The overall review, rating, and the executive feedback will be provided to the evaluation commissioning office.
Parameter Weight (%)	1: Object and context	5	5: Conclusions and lessons learned	20	<i>Are weightings equal to 100%?</i> OK
	2: Purpose and scope	5	6: Recommendations	15	
	3: Methodology	15	7: Gender Equality and Human Rights (UN-SWAP)	10	
	4: Findings	20	8: Presentation	10	

PART I: REPORT DETAILS						
Report title	Corporate Evaluation on Strategic Partnerships for Gender Equality and the Empowerment of Women			Geographical Coverage		Global
Sequence number		Evaluators	[Female]	[Male]	Year	2018
Region	Europe and Central Asia	Country(ies)			Type of intervention evaluated	CPE

Portfolio Budget (USD)		Evaluation Budget (USD)			Reviewer	
Strategic Plan Thematic Area (select all that apply)	Women's leadership and participation	Women's leadership in peace, security and humanitarian response	Women's leadership in peace, security and humanitarian response		Review Date	February 22nd, 2018
	Women's access to economic empowerment and opportunities					

PART II: THE EIGHT KEY PARAMETERS								
SECTION 1: OBJECT AND CONTEXT OF THE EVALUATION (weight 5%)		RATING	Good					
Does the report present a clear and full description of the 'object' of the evaluation?		50%	Executive Feedback on Section 1			Criteria Weight	Weighted increments	Raw point score
1.1 The report clearly specify the object of the evaluation, and provides clear and complete description of the intervention's logic or theory of change, intended beneficiaries by type and by geographic location(s) as well as resources from all sources including humans and budgets, and modalities.		Fully				25%	0.42	1.25
1.2 The context includes factors that have a direct bearing on the object of the evaluation: social, political, economic, demographic, and institutional. This also includes explanation of the contextual gender equality and human rights issues, roles, attitudes and relations.		Mostly				25%	0.42	0.83
1.3 The key stakeholders involved in the implementation, including the implementing agency(s) and partners, other stakeholders and their roles are described.		Partly				25%	0.42	0.42
1.4 The report identifies the implementation status of the object , including its phase of implementation and any significant changes (e.g. plans, strategies, logical frameworks) that have occurred over time and explains the implications of those changes for the evaluation.		Not at all				25%	0.42	0.00
SECTION 2: PURPOSE, OBJECTIVES AND SCOPE (weight 5%)		RATING	Very Good					

Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation?	83%	Executive Feedback on Section 2	Criteria Weight	Weighted increments	Raw point score
2.1 Purpose, objectives and use of evaluation: The evaluation report provides clear explanation of the purpose and the objectives of the evaluation including the intended use and users of the evaluation and how the information will be used.	Fully		50%	0.83	2.50
2.2 Evaluation Scope: The evaluation report provides clear description of the scope of the evaluation, including justification of what the evaluation covers and did not cover (thematically, geographically etc) as well as the reasons for this scope (eg., 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).	Mostly		50%	0.83	1.67
SECTION 3 : METHODOLOGY (weight 15%)	RATING	Very Good			
Is the methodology used for the evaluation clearly described and appropriate, and the rationale for the methodological choice justified?	83%	Executive Feedback on Section 3	Criteria Weight	Weighted increments	Raw point score
3.1 Methodology: The report specifies and provides complete description of a relevant design and sets of methods including the chosen evaluation criteria, questions, and performance standards. The methods employed are appropriate for analyzing gender and rights issues identified in the evaluation scope.	Fully		35%	1.75	5.25
3.2 Data collection, analysis and sampling: The report clearly describes the methods for the data sources, rationale for their selection, data collection and analysis methods. The report includes discussion of how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limitations.	Mostly		40%	2	4.00
3.3 Stakeholders Consultation: The evaluation report gives a complete description of stakeholder's consultation process in the evaluation, including the rationale for selecting the particular level and activities for consultation.	Fully		10%	0.50	1.50
3.4 Limitations: The report presents clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias.	Partly		5%	0.25	0.25
3.5 Ethics: The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards and mechanisms and measures that were implemented to ensure that the evaluation process conformed with relevant ethical standards including but not limited to informed consent of participants, confidentiality and avoidance of harm considerations.	Fully		10%	0.5	1.50
SECTION 4: FINDINGS (weight 20%)	Rating	Good			

Are the findings clearly presented, relevant and based on evidence?	70%	Executive Feedback on Section 4	Criteria Weight	Weighted increments	Raw point score
4.1 The evaluation report findings provide sufficient levels of high quality evidence to systematically address all of the evaluation questions and criteria.	Partly		30%	2.00	2.00
4.2 Findings are clearly supported by and respond to the evidence presented, reflecting systematic and appropriate analysis and interpretation of the data; they are free from subjective judgements made.	Mostly		30%	2.00	4.00
4.3 The causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified.	Fully		20%	1.33	4.00
4.4 Findings are presented with clarity, logic and coherence (e.g., avoid ambiguities).	Fully		20%	1.33	4.00
SECTION 5: CONCLUSIONS AND LESSONS LEARNED (weight 20%)	Rating	Fair			
Are the conclusions clearly presented based on findings and substantiated by evidence?	40%	Executive Feedback on Section 5	Criteria Weight	Weighted increments	Raw point score
5.1 Conclusions are well substantiated by the evidence presented and are logically connected to evaluation findings.	Fully		40%	2.67	8.00
5.2 The conclusions reflect reasonable evaluative judgments that add insight and analysis beyond the findings	Not at all		40%	2.67	0.00
5.3 Conclusions present strengths and weaknesses of the object (policy, programmes, project's or other intervention) being evaluated, based on the evidence presented and taking due account of the views of a diverse cross-section of stakeholders.	Not at all		15%	1.00	0.00
5.4 Lessons Learned: When presented, the lessons learned section stems 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.	Not at all		5%	0.33	0.00
SECTION 6: RECOMMENDATIONS (weight 15%)	Rating	Very Good			
Are the recommendations relevant, useful, and actionable and clearly presented in a priority order?	100%	Executive Feedback on Section 6	Criteria Weight	Weighted increments	Raw point score
6.1 Recommendations are logically derived from the findings and/or conclusions.	Fully		30%	1.50	4.50

6.2 The report describes the process followed in developing the recommendations including consultation with stakeholders.	Fully		20%	1.00	3.00	
6.3 Recommendations are clear, realistic (e.g., reflect an understanding of the subject's potential constraints to follow-up) and actionable.	Fully		30%	1.50	4.50	
6.4 Clear prioritization and/or classification of recommendations to support use.	Fully		20%	1.00	3.00	
SECTION 7: GENDER AND HUMAN RIGHTS (weight 15%)	Score	Approaching Requirements				
Does the evaluation meet UN SWAP evaluation performance indicators? Note: this section will be rated according to UN SWAP standards.	67%	Executive Feedback on Section 7	Criteria Weight	Weighted increments	Raw point score	UN-SWAP score
7.1 GEWE is integrated in the evaluation scope of analysis and evaluation criteria and questions are designed in a way that ensures GEWE related data will be collected.	Partially integrated (1)		33%	1.11	1.11	1
7.2 A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated (2)		33%	1.11	2.22	2
7.3 The evaluation findings, conclusions and recommendation reflect a gender analysis.	Fully integrated (3)		33%	1.11	3.33	3
SECTION 8: THE REPORT PRESENTATION (weight 10%)	Rating	Fair				
Is the report well structured, written in accessible language and well presented?	40%	Executive Feedback on Section 8	Criteria Weight	Weighted increments	Raw point score	
8.1 Report is logically structured, well written and presented with clarity and coherence (e.g. the structure and presentation 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) and written in an accessible language with minimal grammatical, spelling or punctuation errors.	Mostly		40%	1.33	2.67	
8.2 The title page and opening pages provide key basic information on the name of evaluand, 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.	Partly		10%	0.33	0.33	
8.3 The Executive Summary is a stand-alone section that includes an overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key findings, conclusions and recommendations. The Executive summary should be reasonably concise.	Not at all		40%	1.33	0.00	

<p>8.4 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).</p>	Fully		10%	0.33	1.00
Additional Information					
Identify aspects of <i>good practice</i> of the evaluation					

PART III: THE OVERALL RATING			
Key Guiding Question	Total weighted score %	Overall Rating	Other reviewer's comments
Is this a credible report that addresses the evaluation purpose and objectives based on evidence, and that can therefore be used with confidence?	66.83	Good	

