



UNFPA Evaluation Office

Evaluation Quality at UNFPA

Principles and their application



June 2019

Copyright © UNFPA 2019, all rights reserved.

This is an independent document by the UNFPA Evaluation Office.

Any enquiries about this document should be addressed to Ms. Natalie Raaber, raaber@unfpa.org

Table of Contents

1. Introduction	3
2. Background: Evaluation at UNFPA	3
3. The concept of evaluation quality and UNFPA quality assurance and assessment system	4
4. Translating the Principles: Quality Assurance and Assessment in Practice	6

1. Introduction

The UNFPA Evaluation Quality Assurance and Assessment (EQAA) system aims to facilitate the quality of evaluations at UNFPA through two processes: quality assurance and quality assessment. Quality assurance occurs *throughout* an evaluation, beginning with the evaluation's terms of reference and ending with the draft final evaluation report. Quality assessment, on the other hand, takes place after an evaluation is completed (ex post), with the final evaluation report quality assessed by an external independent assessor.

UNFPA Evaluation Office rolled out a revised evaluation quality assurance and assessment system in 2016. Quality criteria used to assess final evaluation reports were updated, and additional quality assurance tools and templates were developed. In 2019, an updated version of Evaluation Handbook: How to Design and Conduct a Country Programme Evaluation at UNFPA was launched, providing additional methodological guidance on integrating gender and human rights in evaluation and evaluating in humanitarian contexts.

The revised EQAA system is aligned with the Evaluation Policy (2019) and the updated UN Norms and Standards for Evaluation (2016). The system covers two types of evaluations: centralized (or corporate) evaluations conducted by the Evaluation Office and decentralized programme level evaluations (namely country and regional programme evaluations). Evaluation quality assessments (EQAs) are posted publically alongside the corresponding evaluation report (and management response) on the Evaluation Database.

Originally authored in 2017, this guidance note was revised in 2019 to include the following:

- additional information on the Evaluation Office's approach to quality assessment, including a description of how the quality assessment scores per criteria are calculated.
- an update to the gender equality and empowerment of women criteria in the EQA grid (criteria #7) to align with the revised UN System-wide Action Plan (UN-SWAP) on Gender Equality and the Empowerment of Women Evaluation Performance Indicator 2.0.

The purpose of this guidance note however remains the same: to further flesh out the concept of (and principles underpinning) evaluation quality at UNFPA, and detail how this is applied/operationalized through assurance and assessment tools. The note is intended to increase the knowledge and understanding of UNFPA evaluation staff, evaluation managers, and external evaluation consultants on quality assurance and assessment at UNFPA and, by so doing, contribute to improved evaluation quality and use for programming and policy decisions.

2. Background: Evaluation at UNFPA

UNFPA adheres to the United Nations Evaluation Group (UNEG) norms and standards for evaluation¹ and its definition of evaluation: "An evaluation is an assessment, conducted as systematically and impartially as possible, of an activity, project, programme strategy, policy, topic, theme, sector, operational area or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency impact and sustainability. An evaluation should provide credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations and lessons into the decision-making process of organisations and stakeholders."

¹ UNEG, Norms and Standards for Evaluation, June 2016

Evaluation at UNFPA serves three main purposes:²

- Evaluation as a means to demonstrate accountability to stakeholders on performance in achieving development results, and on invested resources.
- Evaluation as supporting evidence-based decision making in order to achieve sustainable development results.
- Evaluation as contributing lessons learned to the existing knowledge base on how to accelerate implementation of the Programme of Action of the International Conference on Population and Development and the achievement of the Sustainable Development Goals.

By supporting learning and evidence based decision-making, and demonstrating institutional accountability, evaluation contributes to UNFPA's drive for results. The production of **high quality evaluation reports** is central toward this end; as such the UNFPA evaluation policy and evaluation function place a strong emphasis on **ensuring the quality** of the evaluation exercise and resulting reports, as well as related processes and products.

3. The concept of evaluation quality and UNFPA quality assurance and assessment system

How, then, to understand the concept of high quality evaluations? What are the framing principles and underlying dimensions of evaluation quality?

Overall, UNFPA's evaluation policy adheres to the foundation documents of UNEG and thereby to the norms of utility, credibility, and independence, as well as impartiality, ethics, transparency, human rights and gender equality.³ In 2016, the UNEG Norms and Standards were updated and the approach upgraded to give greater prominence to ethics, transparency and human rights and gender equality.

The UNFPA evaluation policy underscores the importance of quality evaluations to, inter alia, facilitate use.⁴ Accordingly, the Evaluation Office has made efforts in recent years to strengthen evaluation quality, including by developing additional tools and templates to support assurance efforts, and expanding the EQAA system to cover both programme-level evaluations as well evaluations conducted by the Evaluation Office (i.e. corporate /centralized evaluations), which had previously been excluded.

In setting up the evaluation quality assurance and assessment system, three overall and internationally agreed evaluation principles were used:

- **Independence and Impartiality**
- **Credibility**
- **Utility**

Independence and Impartiality

Independence and impartiality are fundamental to the credibility and utility of evaluation. Independence of evaluation entails the ability of evaluators to work without the undue influence of any party. It includes, for instance, freedom for evaluators to choose the design of the evaluation and to select suitable methods and tools for data collection and analysis. Independence also requires that evaluators must not have been directly

² UNFPA Evaluation Policy 2019

³ UNEG, Norms and Standards for Evaluation, June 2016

⁴ UNFPA Evaluation Policy 2019

responsible for the policy setting, design or management of the subject being evaluated. Independence should lead to impartiality, key elements of which are objectivity, professional integrity and absence of bias.

Both independence and impartiality require transparency and should be ensured by the evaluation manager throughout all phases or stages of an evaluation process. Once the evaluation is completed, to uphold these principles, all UNFPA evaluation reports are publicly available on a database maintained by the Evaluation Office.⁵

Credibility

Credibility requires that both independence and impartiality are upheld, but must also be demonstrated and ensured through the competence of the evaluator(s) and the evaluation manager.

The quality of evaluation depends on the design and methodological rigor of the exercise, a transparent and inclusive evaluation process, the reliable use of data, and the ability to perform systematic, logical and sound analysis, reported in a structured way. This, therefore, requires that evaluators hold technical skills on evaluation design, methodology, data collection techniques and a high degree of analytical abilities. A commitment to reflective practice and communication skills are also necessary. Evaluation managers must have the skills to effectively manage the process, including the necessary technical knowledge.

Credibility also relates to the effective integration of human rights and gender equality and ethics in the evaluation process. A high quality evaluation should integrate gender equality throughout the evaluation process - from determining the evaluation scope, criteria and questions through to using a gender responsive methodology, methods and data analysis techniques. The findings, conclusions and recommendations of an evaluation should also reflect a gender analysis.⁶

Both UNEG and UNFPA are committed to ensuring evaluations are gender and human rights responsive. UNEG has, for example, developed guidance on integrating gender and human rights in evaluation,⁷ and UNFPA has directly integrated the UN-SWAP on Gender Equality and the Empowerment of Women Evaluation Performance Indicator criteria among the seven assessment criteria used to assess the quality of evaluation reports, ensuring that gender equality is seen as a critical dimension of evaluation quality. UNFPA also places great importance on ethical principles to be upheld when conducting evaluations, in line with UNEG Ethical Guidelines for Evaluation and the UNEG Code of Conduct for Evaluation.⁸

Utility

Utility requires that evaluations be understood as a tool to influence change and stimulate learning.

For an evaluation to be used, it must be designed from the onset with careful consideration to how the evaluation – in both process and in its resulting products – will impact use. To facilitate this, evaluation managers ought to be skilled in communication, ensuring that reports are concise and easy to read and understand.

To further strengthen utility, an evaluation must be planned/executed with the intention that its results be used. It should be undertaken at a time when its results can meaningfully inform decision-making. It should also be relevant to the needs of intended users, which requires the participation of intended users throughout the evaluation process, including through stakeholder consultations (an important upshot of which is improved ownership).

⁵The UNFPA Evaluation website can be found here: <https://www.unfpa.org/evaluation> and the database of evaluation reports is available here: <https://web2.unfpa.org/public/about/oversight/evaluations/>

⁶ UNEG Guidance Document – UN-SWAP Evaluation Performance Indicator Technical Note, August 2014

⁷ UNEG Guidance Document – Integrating Human Rights and Gender Equality in Evaluations, August 2014

⁸ UNEG Ethical Guidelines for Evaluation and UNEG Code of Conduct for Evaluation in the UN system, June 2008

The principles of utility and use are thus closely linked,⁹ with utilization-focused evaluations in particular enhancing the potential use of evaluations.¹⁰ The UNFPA evaluation policy reflects this principle, stating that evaluations should be “...planned and conducted ensuring national ownership and leadership of evaluation processes by rights holders and duty bearers.”¹¹

4. Translating the Principles: Quality Assurance and Assessment in Practice

The evaluation principles set forth in this guidance form the basis of the quality assurance and assessment system at UNFPA, and ground the evaluation function’s definition of high quality evaluations.

The evaluation principles are translated directly into the tools, templates, and support provided by UNFPA to help facilitate and ensure the quality of evaluation reports. On quality assurance, for example, the Handbook on How to Design and Conduct a Country Programme Evaluation at UNFPA has been revised and updated, templates have been developed to further strengthen the quality of the terms of reference and the design/inception reports, and evaluation teams are vetted to ensure capacity to undertake an evaluation. These quality assurance measures reflect the evaluation function’s commitment to independence/impartiality, credibility and utility.

The principles are further reflected in UNFPA’s approach to quality assessment. The evaluation quality assessment (EQA) grid details seven quality criteria against which UNFPA evaluation reports are assessed. The criteria are further unpacked into sub-criteria, which themselves, if met, signal confidence in an evaluation’s independence/impartiality, credibility and utility. Sub-criteria are weighted differently based on the criteria’s overall importance to the quality of the evaluation report. Previously discussed in a separate document, the approach to weighting sub-criteria is presented below.

Approaching the sub-criteria of the EQA grid

As shared, in 2016, the UNFPA Evaluation Office developed a new evaluation quality assessment (EQA) grid/template. In 2017, adjustments were made, and the approach to the sub-criteria under each quality criteria was clarified to 1) improve consistency among quality assessment reviews and 2) establish objective thresholds for the overall rating scale (unsatisfactory; fair; good; very good).

Under each quality criterion, the relative importance (and weight) of the sub-criteria varies, as detailed in the EQA grid presented below. For the quality assessment of quality criteria 1 to 6, core sub-criteria have been identified in **bold** text. These sub-criteria bear disproportionately on the quality of the evaluation report and, therefore, on the rating of the overall criteria. The absence of any of the core sub-criteria identified for a given quality criterion will result in an “unsatisfactory” rating for the overall quality criterion.

As an example, for quality criterion 1 (*Structure and clarity of reporting*), three core sub-criteria have been identified: sub-criteria 1.1 (*Is the report easy to read and understand?*), sub-criteria 1.3 (*Is the report structured in a logical way, etc.?*) and sub-criteria 5 (*Is an executive summary included in the report etc.?*). If any of these core sub-criteria is assessed as “N” (i.e. sub-criterion not met), the overall rating of quality criterion 1 will be “unsatisfactory”, regardless of the assessment given to the other sub-criteria. If all identified core sub-criteria are either fully or partially met, the overall rating of the quality criterion will range from “fair” to “very good”, based on the combination of values of the other sub-criteria.

⁹ Utilization-Focused Evaluation, 4th Edition, Michael Quinn Patton, 2008

¹⁰ UNFPA Evaluation Policy, 2013

¹¹ UNFPA Evaluation Policy, 2013

A note on the “very good” rating: a “very good” rating should be used to highlight an exemplary level of quality, a good practice and/or a model to follow. As such, all sub-criteria should be met for a quality criterion to be rated as “very good”.

Quality assessment of criterion 7 (*gender equality and the empowerment of women*) directly mirrors the UN-SWAP Evaluation Performance Indication Scoring Tool and is, as such, subject to its associated rating scale.

See the EQA grid below for the specific combinations of values under each quality criterion and the resulting overall rating.

Quality Assessment Criteria	<i>Insert <u>assessment level</u> followed by main <u>comments</u></i>		
<p>1. Structure and Clarity of Reporting</p> <p><i>To ensure the report is comprehensive and user-friendly</i></p> <ol style="list-style-type: none"> 1. Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors? 2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations) 3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)? 4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process? <p><i>Executive summary</i></p> <ol style="list-style-type: none"> 5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation? 6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)? 7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)? 	Yes No Partial	Assessment Level:	
			<p>A rating of “Unsatisfactory” will result if:</p> <p>1,3,5: No (<u>For any of the three</u>)</p> <p>2,4,6,7: Yes/No/Partial</p> <p>A rating of “Good” will result if:</p> <p>1,3,5,6,7: Yes</p> <p>6, 2,4: Partial/Yes</p> <p>7, 2,4: Partial/Yes</p> <p>A rating of “Very Good” will result if:</p> <p>1,2,3,4,5,6,7: Yes</p> <p>A rating of “Fair” will result if:</p> <p>Any combination different from the above.</p>

2. Design and Methodology <i>To ensure that the evaluation is put within its context</i>	Yes No Partial	Assessment Level:	
<ol style="list-style-type: none"> 1. Does the evaluation describe the target audience for the evaluation? 2. Is the development and institutional context of the evaluation clearly described and constraints explained? 3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these? <p><i>To ensure a rigorous design and methodology</i></p> <ol style="list-style-type: none"> 4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection? 5. Are the tools for data collection described and their choice justified? 6. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)? 7. Are the methods for analysis clearly described for all types of data? 8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?) 9. Is the sampling strategy described? 10. Does the methodology enable the collection and analysis of disaggregated data? 11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)? 			<p>A rating of “Unsatisfactory” will result if:</p> <p>2,3,4: No (for any of the three)</p> <p>1,5,6,7,8,9,10,11: Yes/No/Partial</p> <p>A rating of “Good” will result if:</p> <p>2,3,4,5,6,7,9: Yes</p> <p>1,6,8,10,11: Yes/Partial</p> <p>A rating of “Very Good” will result if:</p> <p>1-11: Yes</p> <p>A rating of “Fair” will result if:</p> <p>Any combination different from the above.</p>

3. Reliability of Data <i>To ensure quality of data and robust data collection processes</i>	Yes No Partial	Assessment Level:	
<ol style="list-style-type: none"> 1. Did the evaluation triangulate data collected as appropriate? 2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources? 3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues? 4. Is there evidence that data has been collected with sensitivity to issues of discrimination and other ethical considerations? 		<p>A rating of “Unsatisfactory” will result if:</p> <p>1,2 : No (For any of the two)</p> <p>3,4: Y/P/N</p> <p>A rating of “Good” will result if:</p> <p>1,2,3: Yes</p> <p>4: P</p> <p>A rating of “Very Good” will result if:</p> <p>1-4: Yes</p> <p>A rating of “Fair” will result if:</p> <p>Any combination different from the above</p>	

<p>4. Analysis and Findings</p> <p>To ensure sound analysis and credible findings</p>	<p>Yes No Partial</p>	<p>Assessment Level:</p>	
<p>5. Are the findings substantiated by evidence?</p> <p>6. Is the basis for interpretations carefully described?</p> <p>7. Is the analysis presented against the evaluation questions?</p> <p>8. Is the analysis transparent about the sources and quality of data?</p> <p>9. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?</p> <p>10. Does the analysis show different outcomes for different target groups, as relevant?</p> <p>7. Is the analysis presented against contextual factors?</p> <p>8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?</p>		<p>A rating of “Unsatisfactory” will result if:</p> <p>1,5,7:No(For any of the three)</p> <p>2,3,4,6,8: Y/N/P</p> <p>A rating of “Good” will result if:</p> <p>1,7,8: Y</p> <p>3,2,4,5,6: P/Y</p> <p>A rating of “Very Good” will result if:</p> <p>1-8: Y</p> <p>A rating of “Fair” will result if:</p> <p>Any combination different from the above.</p>	

5. Conclusions <i>To assess the validity of conclusions</i>	Yes No Partial	Assessment Level:	
<ol style="list-style-type: none"> 1. Do the conclusions flow clearly from the findings? 2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated? 3. Do the conclusions appear to convey the evaluators’ unbiased judgement? 		<p>A rating of “Unsatisfactory” will result if:</p> <p>1: No 2,3: Y/P/N</p> <p>A rating of “Good” will result if:</p> <p>1,3: Yes 2: P</p> <p>A rating of “Very Good” will result if:</p> <p>1,2,3: Yes</p> <p>A rating of “Fair” will result if:</p> <p>Any combination different from the above.</p>	

6. Recommendations <i>To ensure the usefulness and clarity of recommendations</i>	Yes No Partial	Assessment Level:	
4. Do recommendations flow logically from conclusions? 5. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)? 6. Do recommendations appear balanced and impartial? 7. Is a timeframe for implementation proposed? 8. Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?		A rating of “Unsatisfactory” will result if: 1,2: No (For any of the two) 3,4,5: Y/P/N A rating of “Good” will result if: 1,2,3,5: Yes 4: P A rating of “Very Good” will result if: 1-5: Yes Fair Any combination different from the above	

<p>7. Gender</p> <p><i>To assess the integration of Gender Equality and Empowerment of Women (GEEW)</i></p> <ol style="list-style-type: none"> 1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected? 2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques? 3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis? 	0-3	Assessment Level:	
		See footnote ¹²	

¹² This assessment criterion is fully based on the UN-SWAP Evaluation Performance Indicator Scoring Tool. Each sub-criterion shall be equally weighted in correlation with the calculation in the tool. The EPI Scoring Tool can be found here: <http://www.unevaluation.org/document/detail/1452>

For more information, please see UNFPA Evaluation Office [webpage](#) where you find the [Handbook](#) for conducting country programme evaluations, [quality assurance](#) templates and checklists, and the new Evaluation Quality Assessment template for assessments of final evaluation reports. All UNFPA programme-level and corporate evaluations are published on the [evaluation database](#), together with the Evaluation Quality Assessment document and a Management Response




Ensuring rights and choices for all

United Nations Population Fund
Evaluation Office
605 Third Avenue
New York, NY 10158 USA

evaluation.office@unfpa.org
www.unfpa.org/evaluation

Follow us on

 [@unfpa_eval](https://twitter.com/unfpa_eval)

 [UNFPA Evaluation Office](#)