

# GEROS Evaluation Quality Assurance Tool

Version: September 2021

## Inter-Agency Humanitarian Evaluation of the response to the crisis in Afghanistan

REPORT RATING SUMMARY			
<b>Overall Rating</b>	63%	Satisfactory	
●●●●●	Exceptional (96% - 100%)	5	
●●●●○	Highly Satisfactory (87.5% - 95.99%)	4	
●●●○●	Satisfactory (62.5% - 87.49%)	3	Meets UNICEF/UNEG standards for evaluation reports and decision makers may use the evaluation with confidence
●●○●●	Fair (35% - 62.49%)	2	
●○○●●	Unsatisfactory (0% - 34.99%)	1	
REPORT DETAILS			
<b>Title of the evaluation report</b>	Inter-Agency Humanitarian Evaluation of the response to the crisis in Afghanistan		
<b>Report sequence number</b>	Evaluation Office 65-2024-20193		
<b>Region</b>	SAR		
<b>Year of report</b>	2024		
<b>Office</b>	Inter-Agency Humanitarian Evaluation Steering Group		
<b>Coverage (countries)</b>	Afghanistan		
<b>ToRs present</b>	Yes		
<b>Date of review (dd/mmm/yyyy)</b>	24-25 June 2024		
<b>Name of review firm</b>	IOD PARC		
CLASSIFICATION OF EVALUATION REPORT			
<b>Management of evaluation (Managerial control and oversight of evaluation)</b>	Jointly managed with one or more UN agencies		
<b>Unicef goal areas (Alignment with strategic plan priorities)</b>			
Every child survives and thrives	Yes		
Every child learns	Yes		
Every child is protected from violence and exploitation	Yes		
Every child lives in a safe and clean environment	Yes		
Every child has an equitable chance in life	Yes		
Gender equality (cross-cutting)	Yes		
Humanitarian action (cross-cutting)	Yes		
<b>Evaluation object</b>	Thematic area		
<b>Evaluation type</b>	Summative		
<b>Evaluation strategy</b>	Mixed methods		
<b>Evaluation design (primary method used)</b>	Participatory		
<b>Evaluation level</b>	Output & Outcome		
<b>Geographic scope</b>	National		
<b>Primary SDG(s) covered (number)</b>	1-6, 16, 17		
<b>EQA Summary:</b> <i>The rater will provide top line issues for this evaluation relevant for feedback to senior management (positive and negative), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight best practice/added value elements and the level of complexity of the evaluation.</i>			
<b>Strengths:</b>			
--The report is well written and contains useful and well-researched insights on the IASC response to the crisis.			
--The evaluation report provides sufficient contextual details to aid in understanding evaluation findings across the different sections.			
--The report is clear in instances where there is a lack of evidence to draw a conclusion.			
--The evaluation report clearly states the evaluation purpose (intended use), scope, objectives and 'why now'.			
--The report offers thoughtful conclusions and detailed and actionable recommendations based on the findings and conclusions.			
<b>Weaknesses:</b>			
--The structure of the report does not allow for readers to easily relate findings to the evaluation questions.			
--The findings do not always clearly indicate the accompanying supporting evidence.			
--Lessons learned do not contain sufficient detail to be useful.			
--Data collection tools and accompanying informed consent processes are not included as annexes.			
--Acronyms are sometimes not defined upon first use.			
<b>Recommendations for Improvement:</b> <i>The rater will identify topline recommendations to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, resources will be cited to assist evaluation managers in overseeing future evaluations.</i>			
Many of the shortcomings highlighted in this review are easily remedied. The scoring scale for the different criteria dictated that some criteria be rated as 'partially' even though the report met most of the criteria's requirements. This has negatively impacted the rating of the report and resulted in a lower overall rating for the report.			
The single major shortcoming in this evaluation report is the inconsistent citations/references. The report should have more systematically cited supporting evidence in the findings section. As written, it is unclear what supporting evidence undergirds many statements/findings.			
The lessons stemming from the evaluation require expansion to be useful.			
Data collection tools and accompanying informed consent processes should be included as annexes.			
SECTION RATINGS			
<b>SECTION A: EXECUTIVE SUMMARY (weight 5%)</b>	67%	Comments on Rating	
<b>Question 1.</b> Can the executive summary inform decision-making?			
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).	Partially	The executive summary is 9 pages in length and exceeds the 2,500 word limit per the ToR (approx. 4,500 words). However, the executive summary reads well, is clearly presented, provides appropriate depth for key users and is appropriately concise given the topic of the evaluation. The executive summary can serve as a standalone document useful for informing decision making.	

	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.	Partially	<p>The ToR does not specify the necessary elements of the executive summary. The ToR simply states that the executive summary is to be 2,500 words in length.</p> <p>That said, the executive summary gives an overview of the nature and causes of the crisis but does not provide a summary of the IASC response. In other words, the executive summary provides overall conclusions on the IASC response but does not provide a summary of the IASC response.</p> <p>The executive summary clearly states the evaluation's purpose. The executive summary does not clearly state the evaluation objectives, intended audience, evaluation methodology (beyond the data sources/methods). The executive summary would have been strengthened if it more clearly outlined the IASC response efforts, objectives, intended audience and broader evaluation methodology.</p>
	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.	Yes	The executive summary does a good job at distilling the main findings, conclusions and recommendations of the evaluation. There is sufficient information to understand the evaluand and the evaluation. There is no new information presented in the executive summary.
<b>SECTION B: BACKGROUND (weight 5%)</b>			57%	Comments on Rating
<b>Question 2.</b>		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.	Partially	The evaluation is clear regarding the evaluand. However, given the topic of the evaluation, this criterion does not quite fit. The evaluation does not provide locations, timelines, cost/budget and implementation status of each intervention. That said, the evaluation does provide an overall budget for the whole humanitarian response and is clear that the 2023 funding has been slashed and that implementation is ongoing. The temporal and geographic scopes are clearly stated.
	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).	Yes	The 'substantive scope' subsection of the report clearly describes the relevant duty bearers and rightsholders. The report highlights the difficulties in assessing the numbers reached as well as relevant disaggregation of these numbers. The report highlights that the definitions/meanings of terms like 'reach' (and 'target achievement') is ambiguous. Figure 4 provides information regarding numbers of persons in need alongside disaggregation by gender, age and disability status. Section 4 assesses the reach and targeting of that scaled-up response.
<b>Question 3.</b>		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.	Yes	The report provides an in-depth analysis of the context leading up to 2021 as well as the period under evaluation (2021 through mid-2023). The report findings, conclusions, lessons and recommendations consider how the context influenced the implementation of the collective action of IASC member organizations to meet the humanitarian needs of people affected by the humanitarian crisis in Afghanistan
	ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	No	Linkages to the SDGs and relevant targets and indicators are not made within the evaluation report or annexes.
	iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	The report indicates the status/needs of the rightsholders throughout the report, including information about persons in need throughout the temporal scope of the evaluation.
<b>Question 4.</b>		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).	Partially	The evaluation findings in the body of the report as well as Annex E identify implementing agencies, right holders, duty bearers and other stakeholders. Linkages - as relevant - were also made, but not via a stakeholder map; instead, these linkages were made within the findings themselves. A stakeholder analysis was required under the ToR, but this analysis is not made available as an annex. It would have been beneficial to annex the stakeholder analysis so that the report contained a clear picture of implementing agencies, right holders and additional duty bearers and other stakeholders alongside their linkages.
	ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	No	Annex A provides funding for Afghanistan by year, by donor (total funding 2021-2023) and the top 12 agency recipients of funding (2021-2023). The specific roles of key stakeholders are discussed in the findings and within the expanded sector analysis annex. However, it would have been beneficial to more clearly indicate the specific contributions and roles of key stakeholders in one section or annex to better lay out the operating environment.
<b>SECTION C: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)</b>			42%	Comments on Rating
<b>Question 5.</b>		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.	Partially	The purpose of the evaluation is clearly defined. Footnote 1 on page 20 clearly answers the 'why it was needed at that point in time' (a requirement of the Humanitarian System-Wide Scale Up Protocols). The report clearly states the intended use of the report in Section 1. The ToR clearly indicates the intended users, but this information is not present in the final report. Paragraph 33 of the final report alludes to the intended users ('With regard to learning, the evaluation takes this to apply to the response in Afghanistan, and the approach of the evaluators centered on encouraging reflection by current aid practitioners, their organizations, and coordination bodies, including the HCT'), but the report would have benefited from more directly stating the intended users. While the purpose of the evaluation is clearly defined (including why it was needed at that point in time) and the report details its intended use, a partial rating must be given as the final report is missing the key intended users.
<b>Question 6.</b>	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).	Yes	The evaluation report is clear that it seeks to achieve and the objectives are the same as those presented in the ToR. In addition, the final report does qualify the objectives based on the way in which the evaluation was conducted in practice (see paragraph 33 of the final report).
	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).	Yes	The evaluation report clearly details the scope of the evaluation and qualifies the purpose of the evaluation (i.e. that the evaluation deals with collective responses and aggregate results by sector and overall, rather than individual programmes and that it was not intended as a substitute for individual agency or programme evaluations). The scope of the evaluation is detailed in Section 1 and is a requirement of the Humanitarian System-Wide Scale Up Protocols. Paragraph 53 of the final report notes some challenges and limitations that affected the scope of the IAHE.
<b>Question 7.</b>	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.	No	The ToR states the IAHE would be theory-drive with the evaluation team creating a ToC for the evaluation. (See paragraph 34 of the ToR: 'This ToR proposes a theory-driven approach to the evaluation. The collective response in Afghanistan currently does not have an explicitly defined Theory of Change. This would need to be developed by the evaluation team at the outset of the evaluation, on the basis of the HRP and consultations with the HCT and other stakeholders, as relevant.'). However, the final report does not include a description of the intended results, parts of the results chain applicable to/being tested by the evaluation or any reference to a ToC developed by the evaluation team at the outset of the evaluation. The final report does not indicate any changes from the ToR.
	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).	No	As mentioned above, the ToR stated that a ToC was to be developed. However, no narrative or graphic of the causal relationship between outputs and outcomes is presented in the final report. The final report does not indicate any changes from the ToR.
	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.	No	As previously stated, the ToR specified that a theory of change was to be created ('This ToR proposes a theory-driven approach to the evaluation. The collective response in Afghanistan currently does not have an explicitly defined Theory of Change. This would need to be developed by the evaluation team at the outset of the evaluation, on the basis of the HRP and consultations with the HCT and other stakeholders, as relevant.').  The final report does not indicate that there were any changes from the ToR and does not provide a ToC.
<b>SECTION D:</b>	<b>EVALUATION DESIGN AND METHODOLOGY (weight 20%)</b>	<b>50%</b>	<b>Comments on Rating</b>
<b>Question 8.</b>	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 and human rights based approach (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.	Partially	The evaluation matrix is available in Annex F. In addition to the questions and sub-questions, the matrix also includes sub-questions, indicators and sources and means of verification. The matrix does not consistently include sub-questions for each question and instead sometimes lists indicators/metrics instead of sub-questions. This is because the evaluation matrix includes 'topics' under the sub-questions column. These topics relate to the 'criteria, indicators, metrics' column and are often not sub-questions at all. It is unclear why sub-questions and topics were combined. It would have been clearer to include the topics as criteria/indicators/metrics and only listed sub-questions in the sub-questions column.  The evaluation also contains an evaluation question that states 'to be confirmed' under D.3: 'Collective accountability [TBC]'. The evaluation report should have included the final evaluation matrix -- the version in Annex F appears to be the draft version.

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.	Partially	<p>The evaluation questions/sub-questions are appropriate for meeting the purpose of the evaluation – 1) to enable learning and provide valuable lessons for future IASC scale-up activations and 2) to ensure accountability of the IASC organizations towards both affected populations and donors.</p> <p>While many evaluation questions/sub-questions also help to meet the objectives listed in the ToR, there were other evaluation questions/sub-questions that did not directly relate to the objectives listed in the ToR. Questions in section A of the evaluation matrix, for example, do not relate directly to the IASC response but rather provide valuable contextual information.</p> <p>There is a disconnect between the evaluation questions and sub-questions presented in the evaluation ToR and the questions in the final report. The report should have included reference to any changes from the ToR. The ToR listed five evaluation questions divided by OECD DAC criteria and cross-cutting issues. The ToR offered 14 sub-questions.</p> <p>The final report evaluation matrix lists 22 evaluation questions (many with their own sub-questions) as well as nearly 60 sub-questions. These questions do not clearly relate back to specific criteria. Instead, the evaluation matrix lists several criteria per overarching topic. For example, section A of the evaluation matrix lists 'Evolution of Afghanistan context, humanitarian priorities, operating context, funding (appropriateness, relevance, coverage)' and then displays the 8 main questions and 16 sub-questions/topics. There is no clear linkage between each evaluation question and OECD DAC or other criteria.</p>
<b>Question 9.</b>	<b>Does the report specify adequate methods for data collection, analysis, and sampling?</b>		
i	Evaluation design and set of methods are relevant and adequately robust for the evaluation's purpose, objectives and scope; and are fully and clearly described.	Partially	<p>The mixed methods approach is relevant to the evaluation's purpose, objectives, and scope. However, more details on the evaluation design and methods are needed to assess their robustness. While Annex G (methodology) describes most methods, additional specifics would have been beneficial. For instance, there's a lack of clarity on how diverse perspectives, including those from vulnerable groups like persons with disabilities and women, were incorporated. The evaluation also does not explain how it aimed for gender balance among respondents. Female participation in KIIs/FGDs consistently lags behind males in each province, and the evaluation report does not highlight this as a potential limitation or outline efforts to achieve gender balance.</p> <p>Additionally, the direct observation/site visits are not sufficiently described. These were supplemented with multiple consultations at each site that are not included in the figures in points i-iii on the first page of Annex G. More details about this activity would have been beneficial to include.</p> <p>The reviewer was unable to review the informed consent process or data collection tools as these were not included as annexes. The reviewer cannot comment on the relevancy or adequacy of the data collection tools as it relates to the evaluation's purpose, objectives and scope. The data collection tools, including the interview protocol, should have been provided as an annex.</p>
ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Yes	<p>Data sources are appropriate and include both qualitative and quantitative sources. The data sources are clearly described in Annex G (methodology). Data sources included key informants from the HCT, staff from UN agencies; Cluster coordinators; representatives from INGOs, national and local NGOs; donor representatives; academics and independent informants; senior DfA officials at the provincial level; and affected people. Further, the data collection activities included document and literature review (including quantitative data), direct observations to a number of sites in urban, peri-urban and rural locations such as health and nutrition centres, water supply facilities, schools, business centres and farms. Annex G also states that a number of consultations with individuals at the site visits took place but are not accounted for in the numbers and the list of key informants annex. It would have been beneficial to include figures on these data collection efforts as well.</p>
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Partially	<p>Annex G provides the detailed evaluation methodology. However, the sampling strategy is not clearly outlined. The report would have benefited from including a clear sampling strategy in Annex G.</p> <p>The annex makes clear that purposive sampling was used to select the key informants (i.e., agency staff currently in post and those who were involved in the response at the time of the Taliban takeover in August 2021), and that a snowball sampling approach was adopted thereafter. The annex also includes the selection criteria for sites across the regions where KIIs and FGDs were undertaken but does not contain selection criteria of KII and FGD participants.</p> <p>The annex does not include how diverse perspectives were captured in the semi-structured KIIs. However, the annex does describe how diverse perspectives were captured in the roundtable discussions by allowing for both in-person and online participation (the latter option allowed women who were working from home because of the Taliban bans to participate).</p>
iv	Clear and complete description of the methods of analysis.	Yes	<p>The methodology (Annex G) states that secondary data was uploaded to a Zotero library and organized for analysis purposes using MaxQDA coding system. Primary data was also analyzed using MaxQDA software (i.e., the detailed interview notes that were coded against an agreed coding tree).</p>
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	No	<p>While the ToR notes that a ToC would need to be developed in the inception period of the evaluation, no narrative or visual representation of a logic model/results framework/ToC was made available as part of the evaluation. Any changes to the proposed methodology in the ToR should have been stated in the final report.</p> <p>The report does not draw causal links between outputs and expected outcomes as no expected outcomes are stated.</p>

	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).	Partially	Paragraph 53 under Section 1 details some of the limiting factors that constrained the evaluation. However, there is no commentary regarding gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators.  The evaluation ToR identified potential risks and suggested mitigation measures. The final report could have been improved by including actual risks and limitations encountered during the evaluation, along with their corresponding mitigation measures.
<b>Question 10.</b>		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 the principles in the 2020 revised UNEG Ethical Guidelines (integrity, accountability, respect, beneficence).	Partially	Annex G details the methodology and contains information regarding the ethical considerations of the evaluation. It would have been beneficial for the body of the final report to include a few lines regarding the ethical considerations under Section 1, Approach and Methods. Annex G explicitly references the UNEG Norms and Standards, Code of Conduct for Evaluation in the UN System, and 2020 Ethical Guidelines for Evaluation. However, the annex does not contain contextualized reference to these. Further, the annex also does not explicitly reference the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or the principles in the 2020 revised UNEG Ethical Guidelines (integrity, accountability, respect, beneficence).
	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.	Partially	Annex G included some discussion of ethical safeguards as they related to data protection and the right to self-determination. Discussion around other appropriate issues relevant to the methodology such as fair representation, compliance with codes for vulnerable groups and avoidance of harm were not included.  The evaluation included a site visit to a high school, but the report makes no reference to the UNICEF procedures for Ethical Research Involving Children. Based on information in the report and annexes, it is unclear whether/the extent to which children were included in the evaluation. The report would have benefitted from clearly outlining the extent to which children were included in the evaluation and reference relevant UNICEF procedures.  The evaluation does not include copies of the consent processes or data collection instruments, so the reviewer is unable to confirm/evaluate whether appropriate ethical safeguards were in place and whether the data collection tools complied with UNEG ethical guidelines.
<b>Question 11.</b>		Does the evaluation incorporate innovative practice that adds value to the evaluation process?		
	i	Innovation practice is used to improve the quality of evaluation process. This could be evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	No	No innovative practices were used in this evaluation.
<b>SECTION E: EVALUATION FINDINGS (weight 25%)</b>			<b>67%</b>	<b>Comments on Rating</b>
<b>Question 12.</b>		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.	Partially	The findings are not organized by evaluation question which prevents the reviewer from determining whether findings systematically address all 22 evaluation questions (many with their own sub-questions) as well as nearly 60 sub-questions and criteria (note that the evaluation questions do not clearly relate back to specific criteria, but the evaluation matrix and section titles list several relevant criteria). An example of an evaluation question that was not answered in the evaluation report is: 'How well have specific needs and vulnerabilities related to gender, age, disability, ethnicity and other factors been addressed?'
	ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	No	The ToR stated: 'This ToR proposes a theory-driven approach to the evaluation. The collective response in Afghanistan currently does not have an explicitly defined Theory of Change. This would need to be developed by the evaluation team at the outset of the evaluation, on the basis of the HRP and consultations with the HCT and other stakeholders, as relevant.'  However, the evaluation report does not make use of a ToC. The report would have benefitted from clarifying that the suggestion in the ToR was not followed, alongside a justification for not developing a ToC.
<b>Question 13.</b>		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.	Partially	Based on the list of interviewees and secondary sources, overall, the evaluation uses credible forms of qualitative and quantitative data. As appropriate, the evaluation expresses concerns regarding accuracy/credibility of certain data points based on other supporting evidence gathered during the evaluation. Triangulation is evident in many findings and both output and outcome-level data are presented. That said, there are many statements/findings that do not include sources/citations for the statements. The report should have included source information as footnotes or within the body of the text for all findings.

	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.	Yes	Both positive and negative findings are clearly supported by, and respond to, the evidence presented (however, as noted above, sources of this evidence are not always cited). The findings are not organized by evaluation question which prevents the reviewer from determining whether findings are based on clear performance indicators, standards, benchmarks or other means of comparison as relevant for each question. Given this determination could not be made and that the criterion states 'as relevant', the reviewer has noted that the report fully meets the criterion to avoid any negative impact on the scoring.
	iii Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings also analyze the logical chain (progression -or not- from implementation to results).	Yes	Causal factors are explored in the findings subsections. The report highlights both enabling and constraining factors that contributed to the achievement or non-achievement of results.  As previously stated, although the ToR proposed a theory-driven approach, the evaluation report does not make use of a ToC and clarification/justification as to why would have been beneficial.
<b>Question 14.</b>	Does the evaluation assess and use the intervention's Results Based Management elements?		
	i Assessment of the adequacy 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) to support decision-making.	Yes	Given the scope of the evaluation, the report appropriately assesses the adequacy of the response's monitoring systems to support decision making. Multiple findings discuss the quality of monitoring and evaluation systems within the response.
<b>SECTION F: EVALUATION CONCLUSIONS &amp; LESSONS LEARNED (weight 10%)</b>		<b>88%</b>	<b>Comments on Rating</b>
<b>Question 15.</b>	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).	Yes	The conclusion section is sufficiently forward looking, well written/clearly formulated and reflect the purpose and objectives of the evaluation. The separation of general conclusions and topic-specific conclusions help readers more easily digest the information presented.
	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.	Yes	The conclusions build upon the findings to provide the evaluation teams' interpretation and analysis of those findings, discussing their implications and significance. The conclusions present both strengths and limitations of the response.
<b>Question 16.</b>	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.	Yes	The lessons shared in paragraphs 270 and 271 stem logically from the findings and have wider applicability and relevance beyond the humanitarian response in Afghanistan (2021-2023).
	ii Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	The lessons are well written and concise but lack sufficient detail to be useful for intended audiences. For example, the following lessons are presented on page 271 but are too vague to be useful: 'With regard to scale-up and transition, lessons from the humanitarian response in Afghanistan include the change needed in the human resource profile, the need for organizational flexibility, and the ability to support a radical shift of strategy and operations.'
<b>SECTION G: RECOMMENDATIONS (weight 15%)</b>		<b>63%</b>	<b>Comments on Rating</b>
<b>Question 17.</b>	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.	Yes	The recommendations align with the purpose of the evaluation (to enable learning for the humanitarian system and to ensure accountability of the IASC organizations towards both affected populations and donors). The recommendations are insightful, clearly formulated and provide essential details for operationalization. The recommendations stem from the conclusions which are derived from the findings.
	ii Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Yes	Recommendations appear useful and are worded in an actionable way for the designated parties to carry forth. Each recommendation includes guidance to aid in implementation including specific steps that should be taken to achieve the larger recommendation.
	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).	No	The final report does not contain any information regarding the process for developing the recommendations.
<b>Question 18.</b>	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.	Partially	The report clearly identifies groups of duty bearers responsible for enacting each recommendation. The report does not provide prioritization or classification of recommendations to support use.
<b>SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)</b>		<b>79%</b>	<b>Comments on Rating</b>
<b>Question 19.</b>	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.	Partially	The opening pages contain 6 of the 7 required components. However, because the report did not include the timeframe of the evaluation in the opening pages, the reviewer must mark this criterion 'partially'.
	ii Annexes include: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework (unless included in report body), 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, information about the evaluator(s), etc.	Partially	The report includes 5 of the 7 required annexes as well as many other appropriate annexes (described in the paragraph below). The report is missing the results chain/ToC/logical framework, which was noted in the ToR but not referenced in the final report, and the data collection instruments.  The report also included thematic annexes (part 1) and evaluation methods (part 2). The thematic annexes detailed the scaled-up response in funding, the Joint Operating Principles, an expanded sector analysis, a Community Consultation report and a thematic summary around the bans on female Afghan staff. The evaluation methodology annexes (part 2) includes the evaluation matrix, methodology, terms of reference, reference documents, list of key informants and the evaluation team itinerary for data collection in Afghanistan.
<b>Question 20.</b>	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).	Yes	The structure is easy to navigate with a table of contents that utilizes hyperlinks for easier navigation. The sections are numbered and the report uses clear titles and subtitles. The report is 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.	Yes	The report structure generally follows the UNICEF guidelines for evaluation report. The context, purpose and brief methodology precedes the findings and the findings are followed by the conclusions, lessons learned and recommendations.
<b>Question 21.</b>	Is the report well presented?		
	i Report is of reasonable length; it does not exceed number of pages that may be specified in ToR.	Partially	The report exceeds the number of words listed in the ToR. The report is quite long - approximately 65 pages without the executive summary. This may be a barrier to readers.
	ii Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	Overall, the report is well written and easy to understand. It is written in an accessible way for the intended audience (per the ToR). There are some instances where acronyms or terminology is not defined upon first use, but given the intended audience, these instances likely did not negatively impact the readers' ability to understand the content. The report is generally free from grammar, spelling and punctuation errors.
	iii Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Yes	The report frequently uses visual aids including text boxes, maps, and figures. These are clearly labelled, presented and referenced in the text.
<b>SECTION I:</b>	<b>EVALUATION PRINCIPLES (weight 10%)</b>	<b>64%</b>	<b>Comments on Rating</b>
<b>Question 22.</b>	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, CCC, CEDAW and/or other rights related benchmarks in the design of the evaluation.	Yes	The final report briefly mentions that the evaluation was guided by the UNEG Norms and Standards, Code of Conduct for Evaluation in the UN System, and 2020 Ethical Guidelines for Evaluation. The ToR states that the 'UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation shall serve as point of reference to integrate human rights and gender equality concepts, standards, values and principles throughout the evaluation'. However, this guidance is not referenced in the final report or accompanying annexes.
	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).	Partially	The evaluation report does not contain a clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation. However, the report an accompanying annexes do give some details regarding the participation of key rights holders and duty bearers. For example, Annex J lists the key informants, so the reader knows what rights holders and duty bearers participated in data collection. Further, Annex G (methodology) states that 'The Evaluation Team also reviewed internal and non-publicly available data shared by the Management Group for this evaluation.' However, there is no further information about the management group's roles and responsibilities, although its members are listed in the opening pages of the report. Given that the report provides a list of key informants (i.e., duty bearers and rights holders were involved as informants) and that the report notes a management group (i.e., a reference group was established) but does not clearly describe the management group's level of participation, this criterion was marked as 'partially'. The report would have benefited from including one or two paragraphs detailing the involvement of key rights holders and duty bearers during the evaluation.

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).	Partially	The report generally utilizes inclusive and empowering language, avoiding gender, heterosexual, age, cultural and religious bias. The report does not use the term 'duty bearers' or 'rights holders'. Instead, the report uses terms like de facto authorities, authorities, IASC members and specific organizations to describe the duty bearers. The report utilizes the term beneficiaries to describe the rights holders.  Where available, data is disaggregated by gender, age and disability. The report highlights differential results across different groups as data allow.
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.	Partially	The evaluation questions and sub-questions included questioning around the extent to which specific needs and vulnerabilities related to gender, age, disability, ethnicity and other factors had been addressed. While the evaluation reflects on the response's inclusion of excluded and marginalized groups such as persons with disabilities, women and children, in-depth analyses of inclusion of these groups were not performed. It is unclear whether this was a result of the methodology or availability of data.
<b>Question 23.</b> Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards with detail provided below)		6	
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.	Fully integrated	The evaluation assesses whether sufficient information was collected during the evaluation period that could help measure progress on human rights and gender equality results. The evaluation featured a specific objective related to gender equality ('Examine the level of gender-responsive programming and women and girls' participation across the collective response'.) Human rights considerations were present in all evaluation objectives. The evaluation framework mainstreamed gender across evaluation criteria/topics. The evaluation matrix featured dedicated questions regarding GEEW (B.2 and E.2).
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Partially integrated	The evaluation employed a mixed-methods approach which is appropriate to evaluating GEWE considerations. The evaluation utilized a range of data sources and processes to help guarantee inclusion, accuracy and credibility (e.g., through triangulation, validation). The evaluation methodology acknowledged and integrated gender considerations and reflections within the analysis and findings and some findings noted how gender influenced certain outcomes. The evaluation methodology used for the roundtable discussion included both in person and online participation – the latter option, as the report noted, allowed women who were working from home because of the Taliban bans to participate.  However, without access to the data collection tools, the reviewer cannot comment on the extent to which the tools were gender-responsive. The evaluation investigated the extent to which available data allowed for disaggregation by sex. Some of the data collected by the evaluation was disaggregated by sex (figure 2 in Annex G provides a breakdown of KIs and FGDs, but no information is available regarding roundtable discussions).  The methodology section did not sufficiently describe how data collection and analysis methods integrated gender considerations. Further, without access to the data collection instruments and associated informed consent process, the reviewer cannot comment on whether ethical standards were considered throughout the evaluation and confirm that all stakeholder groups were treated with integrity and respect for confidentiality. That said, the evaluation report clearly states that UNEG ethical guidelines were followed.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	The report discusses the context in detail which includes reflections on how the humanitarian situation has impacted specific social groups. The findings include data analysis that explicitly and transparently triangulates voices of different social groups. Where available, quantitative data is disaggregated by specific vulnerabilities.  The report does not provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area.

#### SWAP Rating Guidance

<p><b>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.</b></p> <p>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?</p> <p>b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?</p> <p>c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?</p> <p>d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?</p>
<p><b>ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.</b></p> <p>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?</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?</p> <p>d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?</p>

**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 GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?