

GEROS Evaluation Quality Assurance Tool

Version: September 2021

Evaluation of the UNICEF L3 Response in Yemen

REPORT RATING SUMMARY			
Overall Rating		61%	Fair
●●●●●	Exceptional (96% - 100%)	5	
●●●●●	Highly Satisfactory (87.5% - 95.99%)	4	
●●●●●	Satisfactory (62.5% - 87.49%)	3	
●●●●●	Fair (35% - 62.49%)	2	Meets UNICEF/UNEG standards for evaluation reports in some regards, but not all. Decision makers may continue to use the evaluation with caution, but substantive improvements are possible.
●●●●●	Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
Title of the evaluation report	Evaluation of the UNICEF L3 Response in Yemen
Report sequence number	Evaluation Office/65/2024/18946
Region	MENA
Year of report	2024
Office	UNICEF Evaluation Office
Coverage (countries)	Yemen
ToRs present	Yes
Date of review (dd/mmm/yyyy)	October 7, 2024
Name of review firm	IOD PARC

CLASSIFICATION OF EVALUATION REPORT	
Management of evaluation (Managerial control and oversight of evaluation)	UNICEF managed
Unicef goal areas (Alignment with strategic plan priorities)	
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
Evaluation object	System
Evaluation type	Summative
Evaluation strategy	Mixed methods
Evaluation design (primary method used)	Participatory
Evaluation level	Outcome
Geographic scope	National
Primary SDG(s) covered (number)	3, 4, 5, 6, 16, 17

EQA Summary: 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.

Strengths:

- The report is logically structured, well written and is accessible to a wide audience.
- The evaluation report clearly states the evaluation purpose, scope, objectives, 'why now', intended use and intended users.
- The evaluation report provides visual aids throughout the report to help visualise key points.
- The report gathered data from 167 key informants (UNICEF & non-UNICEF) and 104 FGD participants from IRG-controlled and DFA-controlled areas, giving the findings a rich data base.
- The findings are organised by evaluation question and largely answer each piece of each question (with a few exceptions, as noted below). Where data availability or reliability was an issue, the authors clearly indicate this limitation.

Weaknesses:

- The executive summary does not contain sufficient information to be a standalone document useful for decision making.
- The report's length (over 100 pages exclusive of executive summary, Appendix 1 (sector analyses) and annexes). The length of the report may impede uptake of information and reduce accessibility to relevant audiences.
- The report uses the terminology 'beneficiaries' instead of rights-based terminology like 'rights holders' or 'participants' that align with an empowerment-focused approach.
- The final report lacks details about the evaluation reference group's roles and responsibilities.
- The evaluation questions do not always correspond to the criteria referenced in the evaluation matrix.
- Lessons learnt/good practices are not contained in their own section and are instead hidden within findings.
- Data presented in the findings section is sometimes not captured in Appendix 1 (sector analyses), and it is unclear why this is the case. In other words, it is unclear whether the sector analyses were meant to contain all of the detailed data from the evaluation and act as a supplement to the evaluation findings section or whether these two pieces served separate functions – each containing different components of the evaluation data. Clarity on these two pieces would have been beneficial to include in the final report.
- Data collection tools for the FGDs, KIs, direct observations and QuIP activity and accompanying informed consent processes are not included as annexes.
- The report lacks details on how the QuIP activity was carried out and how this data fed into the findings.

Recommendations for Improvement: 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.

- The report would have benefited from a more detailed executive summary. As written, the executive summary does not contain sufficient detail to be a standalone document used to inform decision making.
- The report would have benefited from some revision to cut down on its length. It is currently over 100 pages, exclusive of the executive summary and annexes. This is likely a barrier to uptake/consumption of the evaluation.
- The final report should have referenced ethical considerations and guidelines which are currently missing.
- The report would have benefited from detailing how/if the reconstructed theory of change was utilised for formulation of evaluation findings. The reconstructed theory of change is available in the inception report, but not provided in the final report and should have been included.
- The report would have benefited from a review and revision of evaluation questions and sub-questions and their related criteria. As presented, they do not always correspond to the appropriate criteria in the evaluation matrix.
- The report would have benefited from a brief overview/summary of all efforts included under the evaluation, including roles/responsibilities of key partners/stakeholders for clarity. This could have been organised by sector to match the sector analyses.
- The authors should have included more details on the lessons learned/good practices and included this information as a separate section. As written, they are hidden within the findings, and some do not contain sufficient detail to be useful (as noted in feedback on question 16.i).
- All data collection tools and accompanying informed consent processes should have been included as annexes. More details regarding the QuIP activity should have been provided.

SECTION RATINGS

SECTION A: EXECUTIVE SUMMARY (weight 5%)		50%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
	i Is clearly presented, serves as a standalone document useful for informing decision making, and is of relevant conciseness and depth for key users (Maximum of 5 pages unless otherwise specified in ToR).	Partially	The executive summary is seven pages long. However, considering the breadth and complexity of the evaluation, the length is appropriate. It contains a very high-level overview of the background/context, evaluation purpose/objectives/scope, evaluation methodology and findings per evaluation question. Without additional details on the recommendations, the document does not serve as a standalone document useful for decision making. The executive summary could have removed the photos on pages 1, 3, 5 and 6 to make room for all recommendations (topline and individual recommendations).
	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	The executive summary includes the evaluation purpose, objectives, evaluation methodology and key recommendations. However, it lacks an overview of the intervention (beyond stating UNICEF's programmatic and operational response in Yemen between 2019 and 2022), intended audience, key conclusions and lessons learned. The ToR does not contain a template for the executive summary.
	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.	Partially	The executive summary contains many significant details required to understand the evaluation but lacks depth regarding the intervention itself. The executive summary would have benefited from providing more details on UNICEF's programmatic and operational response between 2019-2022 (e.g., thematic areas, geographic areas, as CLA, etc.). The executive summary does not introduce new information from what is presented in the rest of the report. While there are times where the summarised findings would have benefited from additional details (e.g., "At the same time, challenges in managing staffing resulted in some gaps, notably in terms of sustained capacity and lead times for programme implementation. The evaluation team did not find particular gaps in duty of care responsibilities. Shifting the focus away from recruitment has contributed to a more strategic assessment of the competencies needed in line with the revised country programme document, which prioritizes systems strengthening and values experience in nexus programming. While the outsourcing of administrative functions, including HR-related ones, to the YCO outpost in Amman yielded significant advantages, it created a divide between staff based in Yemen and staff based in Jordan, also in terms of contractual conditions."), the reviewer notes page constraints and the trade-offs required to meet these.
SECTION B: BACKGROUND (weight 5%)		57%	Comments on Rating
Question 2.	Is the object of the evaluation clearly described?		
	i Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status.	Partially	The final report does not provide a clear and relevant description of all interventions included under the evaluation. The report does include information on total funding by UN agency only for 2019, 2022 and 2023 in table 5.
	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/cites, sub-regions...) and in terms of numbers reached, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).	Partially	The report includes numbers around people in need (PIN) per year and IDPs per year, but beyond those numbers, the report does not provide a clear and relevant description of rightsholders by type, by geographic location and in terms of numbers reached with disaggregation by gender, age and disability (although the evaluation notes extreme challenges in monitoring data and data disaggregation). The report notes the overall estimate for the population in Yemen with disabilities, but the it does not note whether these individuals are included in UNICEF's efforts. The report gives some overview of duty bearers, but the report does not do so clearly. The report would have benefited from more clearly stating the populations covered under the UNICEF-supported interventions as well as the specific duty bearers.
Question 3.	Is the context of the intervention clearly described?		
	i Clear and relevant description of the context of the intervention (i.e. relevant policy, socio-economic, political, cultural, power/privilege, institutional, international factors) and how context relates to the implementation of the intervention.	Yes	The report fully addresses this criterion. The document contains a thorough context section that references the CCCs, the political context, institutional context, international factors and relevant policies. The context section also reflects on the socio-economic and cultural contexts in Yemen.
	ii Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	No	The report does not reference the SDGs. There are clear linkages between UNICEF programming and the SDGs, and the report would have been strengthened by detailing these linkages in the evaluation background and overview section.
	iii Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	The context section provides an overview of the status of rights holders of the intervention as they relate to several key thematic areas including health, nutrition, WASH, education, child protection and social protection and cash assistance. This section includes a reflection on availability of data (e.g., what is known vs unknown), which is helpful to frame the evaluation.
Question 4.	Are key stakeholders, their relationships and contributions clearly identified?		

	i	Identification of implementing agency(ies), development partners, right holders, and additional duty bearers and other stakeholders; and of linkages between them (e.g., stakeholder map) (if relevant).	Partially	<p>The evaluation report offers an overview of broad stakeholders in table format in the section titled 'evaluation stakeholders and intended users' but does not specifically name the relevant ministries, NGOs and donors as well as other relevant UN agencies.</p> <p>The report does not provide information on linkages between implementing agency(ies), development partners, rights holders, and additional duty bearers and other stakeholders.</p> <p>The report would have benefited from including a stakeholder map so that the readers were aware of the full scope of partnerships and coordination structures covered under the evaluation alongside the stakeholders engaged in each.</p> <p>Since the evaluation report identifies some of the key stakeholders, this criterion is marked 'partially'.</p>
	ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	<p>The report does not identify specific contributions and roles of key stakeholders except for table 5 which details the financial contributions of UN agencies only for 2019, 2022 and 2023. The report would have benefited from including clear details regarding contributions of all actors so that readers could understand the full scope of the evaluation.</p> <p>The evaluation includes a very high-level overview of the situation in Yemen with regard to humanitarian actors in the subsection titled 'The international humanitarian response'. However, this subsection does not contain details that satisfy this criterion. For example, it does not specify which clusters UNICEF leads vs provides support to other CLAs, does not indicate which ministries are partners, etc.</p>
SECTION C: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)			58%	Comments on Rating
Question 5.		Is the purpose of the evaluation clearly described?		
	i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	The purpose of the evaluation and the intended users and uses are described on pages 16-17. This also includes the reason for this evaluation.
Question 6.		Are the objectives and scope of the evaluation clear and realistic?		
	i	Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR (if applicable).	Yes	The objective and purpose of the evaluation is clearly stated on page 16 and references those from the ToR (i.e., no changes from the ToR).
	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	<p>The evaluation report clearly states what is within scope (page 18). Importantly, the report specifies that, geographically, the evaluation scope is the whole of Yemen, 'as relevant with regard to UNICEF's strategy and programming in the country, as per its six field offices'.</p> <p>As noted above, the report would have benefited from including a table or visual of all UNICEF programming included under the evaluation. As written, it is difficult for the reader to see the full picture of the evaluation.</p>
Question 7.		Is the theory of change, results chain or logic well articulated?		
	i	Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	No	<p>The reconstructed theory of change is well articulated in the inception report, but is absent from the evaluation report and its accompanying annexes. The final evaluation report does not provide a clear description of the intended results and the parts of the ToC that are being tested by the evaluation.</p> <p>Without information on the theory of change, it is difficult for readers to understand how some evaluation analysis was carried out – especially regarding the QuIP analysis which stated: 'This information is gathered directly from targeted populations, based on their perceptions of what has changed in their lives over a set period of time and across a series of domains related to a given theory of change. By working backwards, from outcomes to what respondents cite (unprompted) as the main drivers of change, QuIP tries to reduce 'confirmation bias,' the risk that people will say what they believe evaluators may want to hear.'</p>
	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	<p>Causal relationships between outputs and outcomes are not presented in narrative or graphic form. A logic model is presented in Figure 1, but this displays the 'evaluation benchmarks', 'analytical dimensions for the evaluation of UNICEF's Yemen Response' and 'evaluation criteria'. Figure one notes a 'reconstructed Theory of Change', but no further information is available in the final report.</p> <p>Section 4.3 of the inception report clearly states that 'no explicit Theory of Change [was] in place that holistically covers the rationale for UNICEF's efforts in Yemen across the extensions.'</p> <p>The evaluation reconstructed a ToC and presents this in the inception report, but no reference to the reconstructed ToC is made in the final report aside from it being listed as an 'evaluation benchmark' in Figure 1. This information should have been included.</p>
	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.	Partially	<p>The reviewer has marked this as 'partially' because the evaluation team determined during the inception phase that no explicit ToC existed that covered the rationale for UNICEF's efforts in Yemen across the extensions. The evaluation team reconstructed a ToC during the inception phase and this is available in the inception report.</p> <p>However, there is no reference to this in the final report aside from noting the</p>
SECTION D: EVALUATION DESIGN AND METHODOLOGY (weight 20%)			59%	Comments on Rating

Question 8.	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? <i>UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender 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	<p>The evaluation questions allow the evaluation to meet its objectives of identifying what UNICEF has done well, what processes have supported the YCO in implementing the response and identifying key gaps as well as lessons learned. The evaluation questions are divided into four broad categories which create some confusion as it appears that some questions are asking similar things and could have been combined. It is unclear how/why this structure was adopted, and an explanatory note/justification would have been beneficial to include. The four categories were: UNICEF as an operational actor (as listed in the evaluation matrix but changed to UNICEF as a humanitarian actor in the final report), UNICEF as an institution, UNICEF as a partner and UNICEF as a Cluster Lead Agency. Note that the findings section incorporates EQ 3.2 into EQ 1 (UNICEF as a humanitarian actor) which was originally listed under UNICEF as an institution in the evaluation matrix. It is unclear why this change was made.</p> <p>The relevant criteria are specified in the evaluation matrix and listed under the evaluation question in the findings section. However, the referenced criteria are not always aligned with the listed evaluation questions.</p> <p>There are several evaluation sub-questions/sub-questions that do not clearly link to the referenced (i.e., coinciding) criteria. Examples which illustrate the need for better alignment between the listed evaluation questions and evaluation criteria include sub-question 1.7 which relates more to 'relevance/appropriateness' than 'effectiveness', sub-question 3.7 which relates to 'efficiency' rather than 'relevance/appropriateness' and sub-question 3.8, which relates more to 'coherence' than 'effectiveness'.</p>
	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.	Yes	In addition to the questions and sub-questions, the evaluation matrix includes the OECD DAC criteria, indicators, data sources and methods for data collection.
Question 9.	Does the report specify adequate methods for data collection, analysis, and sampling?		
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.	Yes	The final report sufficiently details the evaluation design and methods used under the evaluation. The mixed-methods approach and data collection techniques were adequately robust for the evaluation's purpose, objectives and scope.
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	Data sources are appropriate and described in the 'methodological approach' subsection and detailed in the annexes. The data sources include a range of stakeholders through qualitative and quantitative engagements. The data sources also include secondary reviews of quantitative and qualitative data.
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>The final report provides information on the types of sampling approaches for KIIs (purposive and snowballing with both UNICEF and non-UNICEF stakeholders at country, regional and HQ levels), FGDs (project sites chosen through stratified sampling approach to ensure inclusion of different groups and a variety of geographic locations), direct observations (carried out alongside the FGDs) and QuIP approach (purposive sampling, but note that details on the sampled population are not provided in the report). The sampling approach for the partner online survey was not provided and should have been noted.</p> <p>The final report notes how the data gathering activities of KIIs and FGDs ensured that diverse perspectives were captured. The final report does not note how diverse perspectives were captured under direct observations, QuIP approach and the partner online survey. The final report should have included clearer language on how diverse perspectives were captured for each data gathering method.</p>
iv	Clear and complete description of the methods of analysis.	Partially	The methodological approach subsection and figure 2 partially describe the methods of analysis, but do not state how FGD data, direct observation data and survey data was analysed. This should have been included (e.g., thematic analysis, descriptive analysis, etc.). The final report does provide information on how KII data were analysed (i.e., using MAXQDA, a qualitative data analysis software). The final report also provides detail on the QuIP study including the structure of how information is gathered and that the QuIP analysis was performed using Causal Map software developed by Bath SDR.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Partially	The methodology allows for drawing causal connections between outputs and expected outcomes, but these outputs and outcomes are not explicitly stated in the final report. The theory of change is only available in the inception report and is not noted in the final report.
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).	Yes	Table 13 clearly details the limitations faced by the evaluation alongside mitigation strategies to address these limitations.

Question 10.	Are ethical issues and considerations described? The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:		
	i Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or the principles in the 2020 revised UNEG Ethical Guidelines (integrity, accountability, respect, beneficence).	No	The final report does not contain information around the UNEG obligations of evaluators or the principles in the 2020 revised UNEG Ethical Guidelines. These should have been referenced in the final report.
	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.	No	The final report does not contain information around ethical safeguards for participants. This should have been included in the final report.
Question 11.	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.	Partially	The evaluation team used an innovative methodology called the Qualitative Impact Assessment Protocol (QulIP) to gauge impact. However, the final report is lacking details to determine how innovative of an approach this method was. Additional details of the areas of investigation, the sampled population, the analysis process in developing 'causal maps', etc. could have been stated in the body of the report or as an annex.
SECTION E:	EVALUATION FINDINGS (weight 25%)	67%	Comments on Rating
Question 12.	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	Findings offer rich insights and it is clear that a massive amount of data gathering, data reviewing, and data analysis has taken place. The findings section provides a succinct analysis per evaluation question/sub-question with additional details available in the sector analyses (Appendix 1). The findings do not always marshal sufficient levels of evidence to systematically address all of the questions (e.g., EQ 1.1 is not sufficiently answered with regard to appropriateness of strategies, preparedness and response plans). The evaluation team should have indicated when an evaluation question could not be answered, e.g., due to lack of data. Findings often clearly triangulate statements, calling in data from both quantitative and qualitative sources. However, there are many findings that lack reference to supporting evidence. For example, under Programme Objectives (p. 48): 'UNICEF invariably undertook the bulk of WASH activities against the targets set by the WASH cluster. Service providers for water and wastewater supplies continued functioning because of UNICEF support in the most difficult operating context, even at the height of the political and security crisis.' The findings section does not provide reference to supporting evidence in this case. However, if the reader consults the nutrition sector analysis in Appendix 1, they will see that the finding comes from an analysis of top line figures from HRPs and HACs (table 1 in Appendix 1). The evaluation report should have more clearly explained the purpose of the sector analyses and provided a rationale for why some data is presented only in the findings section and not included in the sector analyses.
	ii Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Partially	While it is evident that extensive research and analysis informed the findings section and sector analyses (Appendix 1), it is unclear how the reconstructed theory of change was applied in formulating the findings. The findings section includes statements like 'UNICEF largely achieved its intended objectives'. However, it does not clearly outline what those specific objectives were or present them alongside the results evaluated by the team, which led to the conclusion that certain objectives were partially or fully achieved. While Appendix 1 (Sectoral Analyses) provides detailed information on objectives by sector, these objectives do not align with those outlined in the reconstructed theory of change. The findings section does not clearly utilise the intervention's results framework/reconstructed ToC in the formulation of findings. Although the logic model for the evaluation identifies the reconstructed ToC as an evaluation benchmark (i.e., 'stated programme/project/activity objectives as provided in the reconstructed theory of change'), the evaluation instead uses a different set of objectives in the sector analyses (Appendix 1), which supplement the findings section. As an example, the objectives listed for the health sector in the sector analysis (Appendix 1) and the reconstructed theory of change (presented in the inception report) do not match. The differences between the objectives listed in the reconstructed theory of change (and noted as an evaluation benchmark in the logic model for the evaluation) are present for all sectors. The sector analyses appear to respond to specific intervention results frameworks and the findings as presented in the findings section reflect this analysis. As such, the reconstructed ToC does not appear to be used in the formulation or presentation of findings. It is unclear why the reconstructed ToC was listed in the logic model for the evaluation and the value of the reconstructed ToC is questioned.
Question 13.	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.	Yes	<p>The evaluation uses available secondary quantitative data but notes issues and major challenges around accuracy and reliability. The evaluation used credible forms of qualitative data from a range of diverse stakeholders with knowledge of UNICEF's interventions.</p> <p>The sector analyses in Appendix 1 presents output and outcome information with accompanying data where available. The findings section then summarise the data from the sector analysis.</p> <p>Triangulation is often evident through the use of multiple data sources. There are times when triangulation was not performed and this is likely due to the lack of data.</p>
	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.	Partially	<p>The evaluation findings include both positive and negative aspects of UNICEF's efforts during the L3 response. Many of the findings are clearly supported by evidence either referenced in the text or provided as footnotes. However – and as noted above – the evaluation findings section also contains many findings that lack reference to sources for the information presented. The findings section does not always link the findings to clear performance indicators, standards or benchmarks. However, the sector analyses (Appendix 1) clearly link the findings to specific objectives or indicators for each intervention. The final evaluation report would have been strengthened if the authors performed a thorough review of all evaluation findings and ensured that supporting evidence was clearly indicated either as a source citation or within the body of the text.</p>
	iii	Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings also analyse the logical chain (progression -or not- from implementation to results).	Yes	<p>To the extent that data allowed, the evaluation findings clearly indicate the causal factors leading to achievement or non-achievement of results (e.g., non-achievement: school closures due to COVID, delays in procurement, delayed technical assessment leading to the reprogramming of suspended grants, delays at ports; achievement: interest in activities by the community, functioning latrines in schools). As noted previously, it is difficult to discern how the logic chain was used during data analysis.</p>
Question 14. 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.	Partially	<p>The report contains a specific subsection on programme monitoring, reporting and quality assurance which coincides with evaluation question (EQ) 3.2: 'Were the UNICEF programme monitoring, reporting and quality assurance systems appropriate for the response?' A review of information presented under 3.2 reveals that the assessment of programme monitoring is incomplete.</p> <p>This section includes an analysis of UNICEF's approaches to monitoring (i.e., third-party monitoring, Integrated Supportive Supervision (ISS) model). However, the section lacks reference to other monitoring tools such as the HPM, HAC indicators, the Monitoring and Reporting Mechanism (MRM) for Child Protection, the media monitoring programme, the Nutrition Cluster Performance Monitoring report, among others, which are referred to under other EQ findings (e.g., EQ 1.1). The findings under EQ 3.2 should have included a complete assessment of all monitoring tools, which is currently lacking.</p> <p>That said, the findings under EQ 3.2 highlight critical gaps in the monitoring process. However, the evaluation does not include an assessment of the completeness or appropriateness of the results framework or a discussion of vertical and horizontal logic. This is likely due to the fact that the overall response did not have a ToC. Nevertheless, it is unclear why the evaluation did not make use of the reconstructed ToC developed by the evaluation team during the inception phase.</p> <p>The report does not include an assessment of how/whether the specific M&E tools supported decision-making. This level of detail would have provided rich insights for lessons learned and recommendations and would have responded to some of the objectives of the evaluation (e.g., what UNICEF has done well and what processes have supported the YCO in implementing the response, key gaps).</p>
SECTION F: EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)			75%	Comments on Rating
Question 15. 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	<p>The conclusions section is clearly written and touches on key areas relevant to the evaluation objectives, including identifying what UNICEF has done well and the processes supporting the YCO's response.</p>
	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	<p>The conclusions section provides insights and judgements based on the findings presented in the findings section and elaborated on in Appendix 1 (sector analyses). The section provides concise judgments on the intervention's overall successes and shortcomings.</p>
Question 16. 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.	Partially	<p>Lessons learned are presented within findings (e.g., under EQ 3.4., EQ 3.7, EQ 3.8, EQ 4) which makes it difficult for the reader to pull out these lessons. The lessons stem logically from the findings but it is not always clear whether the lessons have wider applicability and relevance beyond the object of the evaluation due to insufficient details.</p> <p>--The lessons highlighted under EQ 3.4 on the need for coordination across the organization to support the response, the issue of customization versus standardization, and the tension between response and preparedness stem logically from the findings and are sufficiently detailed.</p> <p>--The lesson under EQ 3.7 on building a comprehensive risk-informed institutional framework stemmed from the findings and was presented with sufficient detail.</p> <p>--The lesson under EQ 3.8 on the PMU use of the MIS and the Financial Service Provider Model to deliver social assistance stemmed from findings and was presented with sufficient detail.</p> <p>--The lesson under EQ 4 regarding the Emergency Cash Transfer/Unconditional Cash Transfer programme experience of contracting financial service providers does not contain enough information to clearly identify a lesson learned.</p> <p>The report would have benefited from highlighting each lesson either as call-out boxes within the findings section or within a separate lessons learned section.</p>
	ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	<p>As noted above, lessons learned are presented within findings (e.g., under EQ 3.4., EQ 3.7, EQ 4) which makes it difficult for the reader to pull out these lessons. Some lessons are presented clearly and concisely with sufficient details. Other times, the lessons lack the details to be useful to the intended audience (examples shown above).</p> <p>Per feedback above, the report would have benefited from highlighting each lesson either as call-out boxes within the findings section or within a separate lessons learned section.</p>
SECTION G: RECOMMENDATIONS (weight 15%)			50%	Comments on Rating
Question 17.		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	<p>The recommendations align with the evaluation's purpose of learning and accountability. They are logically derived from the findings; each recommendation is supported by topline findings in the first column of Table 10. Table of Recommendations.</p> <p>All recommendations are clearly formulated.</p>
	ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	<p>Overall, recommendations are sufficiently detailed, useful and appear actionable for the primary intended users. Guidance is given on implementation, as appropriate.</p> <p>However, there are two recommendations that would have benefited from additional information:</p> <p>--Recommendation 2.1 would have benefited from more details regarding what is meant by 'support country offices': 'Support country offices to ensure programme coherence and programme integration to avoid 'silos approach' due to donor fragmentation and increasing earmarking.'</p> <p>--Recommendation 7.3 would have benefited from more details regarding what is meant by 'leverage' and 'strengthen': 'Leverage UNICEF's programmatic and operational role in Yemen to strengthen its leadership role in inter-agency initiatives (e.g., on access, systems strengthening).'</p>
	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 report does not contain information on the process for developing the recommendations and this should have been included.
Question 18.		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 recommendations clearly identify the groups responsible for enacting each recommendation. However, the recommendations do not offer any prioritisation or classification of recommendations to support use. To strengthen the recommendations section, the evaluation team could have prioritised or provided some recommended timelines for each recommendation.
SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)			71%	Comments on Rating
Question 19.		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.	Yes	The report fully meets this criterion. The opening pages include the name of the evaluated object, location of the evaluated object, the date of the report, name of the organization commissioning the evaluation, the names of the organization of evaluators (in the preface), table of contents with page numbers, and list of acronyms.
	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	<p>Annexes include the terms of reference, evaluation matrix, list of interviewees and their organisation's location, list of focus group discussion locations and participants (i.e., list of site visits), list of documents reviewed and the light partner survey.</p> <p>The report is missing annexes that contain the reconstructed ToC and the rest of the data collection instruments (e.g., direct observation checklist, KII guides, QuIP analysis tool).</p>

Question 20.	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 report is well formatted and easy to navigate. The table of contents contains hyperlinks for easy navigation directly to a section from the table of contents. The formatting allows readers to follow along without the use of numbered sections due to its clear structure and clear titles/subtitles.
	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 final report structure accords to UNICEF guidelines for evaluation reports.
Question 21.	Is the report well presented?		
	i Report is of reasonable length; it does not exceed number of pages that may be specified in ToR.	No	The report is quite long, totalling nearly 100 pages exclusive of the executive summary and annexes. There is no page limit specified in the ToR. However, it is good practice to keep final evaluation reports to 60 pages maximum, excluding annexes. This limit ensures that the report remains concise, focused, and accessible while still providing the necessary detail. The subject of the evaluation is complex, but decreasing the report length could have increased utility and readership. As written, the length of the report presents a major barrier to uptake, especially considering the availabilities/bandwidths of the intended audiences.
	ii Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	The report is easy to understand and general free from grammar, spelling and punctuation errors. A final copyedit/proofread may have caught the remaining grammatical/punctuation errors, but these are few and far between and do not impact the reader's ability to consume the information.
	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.	Partially	The report makes frequent use of visual aids which convey key information and add richness to the report. With the exception of the visual 'six core cluster functions (+accountability to affected people)' on page 95, all of the visuals are clearly labelled. Nearly all of the visuals are referred to in the text, either by discussing the information in the accompanying visuals or by noting the visual title. Examples of visuals that are presented in the report but not specifically referenced by title in the text include figures 4-7. There is also one instance where a figure is referred to incorrectly. A final copyedit would have likely caught this inconsistency: Under timeliness and due diligence (EQ 4.2 and EQ 1.5), the findings refer to Figure 20 when it should have referred to Figure 23. Further, there is a list of figures and list of tables following the acronyms list for easy reference.
SECTION I:	EVALUATION PRINCIPLES (weight 10%)	60%	Comments on Rating
Question 22.	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 evaluation references and utilises the CCC in the design and evaluation of the programme.
	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 report clearly indicates which stakeholders were involved as informants during data collection but does not include details regarding the roles and responsibilities of the evaluation reference group (ERG). The ERG was noted in the preface, but no other information is available in the final report. Similarly, very few details about the ERG are available in the inception report. (Text noted below for easy reference.) The evaluation would have benefited from more fully describing the role of the ERG. ERG reference in the final report preface: 'A special thanks is due to colleagues on the Evaluation Reference Group and the UNICEF Yemen Country Office who contributed valuable time to the evaluation process, including the fieldwork, and through the various iterations of the evaluation report.' ERG reference in the inception report: 'For the data collection process to be thorough and efficient, the evaluation team will work together with the Evaluation Reference Group to ensure that relevant key informants are appropriately identified.'
	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 uses language that is empowering and inclusive (e.g., persons with disabilities and not disabled persons) and avoids gender, heterosexual, age, cultural and religious bias. However, the report does not use the terminology of rights holders and duty bearers. Instead, the report references beneficiaries. The report notes shortcomings regarding the availability of disaggregated data which has 'challenged UNICEF's ability to comprehensively reach the most vulnerable groups'. The report does note some data disaggregated by sex and minority groups (e.g., the 2021 study carried out by the Yemen Humanitarian Fund and Yemen Protection Cluster; see footnote 73) as well as summary targets that are available by age (e.g., under 5 in figure 11). With the exception of KIIs, data gathered under the evaluation (e.g., survey results, FGD participants) are not disaggregated by gender and this disaggregation should have been provided.

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	<p>The evaluation centres around child rights. The evaluation also reflects on the extent to which Leave No One Behind has been implemented during the timeframe of the evaluation. For example, the evaluation assesses the extent to which UNICEF was able to reach the most vulnerable groups in evaluation question (EQ) 2. Additionally, EQ 3.2 assessed the extent to which monitoring took into consideration inclusive dimensions such as persons with disabilities and minority groups, concluding that efforts have been inconsistent. EQ 3.3 reflected on the staffing difficulties and their effects on inclusion, notably leading to a 'more limited integration of social protection considerations/priorities across UNICEF programmes as a whole, notably with regard to more marginalized groups such as children with disabilities'. Lastly, the report highlights under 'programme objectives' findings that targeted programming for vulnerable groups was not clearly spelled out.</p> <p>Ultimately, the evaluation team determined that approaches to engage and support excluded and marginalised groups (e.g., persons with disabilities, Muhamasheen minority group) were limited in scale. Further, the evaluation acknowledged that the 'contextual constraints, the lack of disaggregated data at different levels and challenges with direct programme oversight and programme integration have challenged UNICEF's ability to both comprehensively reach the most vulnerable groups and know who was not reached'.</p> <p>The evaluation itself is partially disability inclusive. The report provides context for persons with disabilities living in the country and includes some analysis of findings related to persons with disabilities. The methodology in the inception report notes that the sampling approach for FGDs was chosen to ensure the appropriate inclusion of different groups, including persons with disabilities, but beyond that declaration, the report contains no information on how persons with disabilities were included as informants for the evaluation. Further, the report contains no information on ensuring appropriateness of data collection tools for persons with disabilities and does not disaggregate FGD participants or KIIL respondents by disability status.</p>
Question 23.	Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards with detail provided below)	5	
	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.	Satisfactorily integrated	<p>SWAP rating guidance: 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.</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? YES.</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? MAINSTREAMED.</p> <p>c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? MAINSTREAMED.</p> <p>d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? SOMEWHAT INTEGRATED.</p>
	ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	<p>SWAP rating guidance: A gender-responsive methodology, methods and tools, and data analysis techniques are selected.</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? PARTIALLY. The data analysis integrates gender considerations but data collected during the evaluation is not disaggregated by sex.</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations? YES.</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? YES.</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? YES. However, the extent to which the stratified sampling approach to project sites ensured the appropriate inclusion of different groups (IDPs, non-IDPs, Muhamasheen, women, men, elderly, people with disabilities) is unknown since data collected during the evaluation is not disaggregated to this level.</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? YES. However, the ethical standards and considerations were not summarised in the final report (only available in the inception report).</p>

iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	<p>SWAP rating guidance: The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.</p> <p>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? PARTIALLY. The final report notes the CCCs, but does not highlight other instruments or policies related to human rights and gender equality (e.g., CRC, CEDAW, CRPD).</p> <p>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? PARTIALLY. The findings triangulate voices of different stakeholders, but the findings section does not specify which groups specifically for specific findings. The findings do, however, provide disaggregated quantitative data as availability allows.</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? NO. Unanticipated/unintended effects of the interventions are not discussed.</p> <p>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? YES. See recommendations 1.3, 4.2, 4.3, 5.1, 5.2, 5.3, 5.5, and 5.7.□</p>
-----	---	----------------------	--

SWAP Rating Guidance

<p>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.</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>ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.</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>
<p>iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.</p> <p>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?</p> <p>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?</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described?</p> <p>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?</p>