

GEROS Evaluation Quality Assurance Tool

Version: September 2021

Iterative evaluation of the UNHCR/UNICEF blueprint for joint action for refugee children

REPORT RATING SUMMARY

Overall Rating	73%	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. Decision makers may use the evaluation with confidence
●● Fair (35% - 62.49%)	2	
● Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS

Title of the evaluation report	Iterative evaluation of the UNHCR/UNICEF blueprint for joint action for refugee children
Report sequence number	Evaluation Office-65-2023-19882
Region	HQ
Year of report	2023
Office	Various
Coverage (countries)	Bangladesh, Cameroon, Ecuador, Ethiopia, Honduras, Indonesia, Iraq, Lebanon, Libya, Rwanda, Uganda and Italy
ToRs present	Yes
Date of review (dd/mm/yyyy)	February 16, 2024
Name of review firm	IOD PARC

CLASSIFICATION OF EVALUATION REPORT

Management of evaluation (Managerial control and oversight of evaluation)	Jointly managed with one or more UN agencies
Unicef goal areas (Alignment with strategic plan priorities)	
Every child survives and thrives	Yes
Every child learns	No
Every child is protected from violence and exploitation	No
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	Strategy
Evaluation type	Summative and formative
Evaluation strategy	Mixed methods
Evaluation design (primary method used)	Participatory
Evaluation level	Output & Outcome
Geographic scope	Multi-region/Global
Primary SDG(s) covered (number)	4, 5.3, 6, 8.7, 16.2, 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.

This is a good quality report in many ways that should provide UNICEF with useful insights and recommendations for a future partnership with UNHCR in this area. It is missing some elements of analysis proposed in the design, however, that should (or should have) been addressed. Some key strengths and weaknesses are as follows:

- The Executive Summary contains all relevant information but is overly lengthy and is more of a selection of key passages from the main report rather than a genuine summary.
- There is a comprehensive overview of the partnership, including its vision, outcomes, and timeline, but lacks some key details, such as the initiative results framework, an explanation of "Blueprint 2.0", description of the situation of rights-holders and duty-bearers, particularly at the country level, financial contributions
- The evaluation questions and scope align well with the overall objectives, but could benefit from clearer differentiation between data collection and analysis, as well as inclusion of data collection tools in the annexes.
- The report briefly discusses innovative practices noted in the Inception Report but does not clarify if they were utilized in the final analysis, an aspect requiring clarification.
- The report provides a thorough explanation of data analysis methods, detailing how primary and secondary data were treated and analyzed, including triangulation and stakeholder participation in validation.
- The structure of the report is generally clear, with explanatory text preceding each subsection, generally aligning with the evaluation matrix (although some elements have been combined), although findings lack itemization or numbering.
- Findings are useful and present available data well to support analyses, providing insights into key challenges and adaptations, aligning with recommendations for improvement.
- The Conclusions offer a comprehensive summary of both positive and negative aspects of the initiative, structured according to evaluation questions, but do not reference specific findings.
- Lessons learned are interspersed throughout the report, although a standalone section would have been preferable to fully capture and detail them, particularly given the centrality of this aspect to the evaluation objectives
- Recommendations cover fundamental, management, and relationship aspects of partnerships, although dense text could be better presented as bulleted action points and lack some prioritization, responsible entities, and information on time horizons.
- The annexes provide additional information, but the report could benefit from inclusion of data collection tools and additional details on initiative implementation.
- The report's style and writing are generally good, but inconsistencies in spelling, formatting, and numbering detract from its quality.
- Gender and equity considerations are mentioned in the methodology but lack further analysis or unpacking in the scope or 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.

Key areas for improvement are as follows:

- Shorten the Executive Summary, condensing introductory/contextual information, methods/design, and findings into bulleted points, while summarizing recommendations into a single page with actionable bulleted action points.
- Clarify the status and needs of rights-holders benefiting from the initiative, potentially including more detailed information in an annex to ground the evaluation report in concrete outcomes.
- Improve clarity regarding the role of the evaluation vis-a-vis 'Blueprint 2.0,' including clear explanations of timing, duration, and contributions to this phase.
- Address inconsistencies between the stated objectives in the evaluation report and those outlined in the Terms of Reference (TOR), ensuring alignment and clarity throughout the document.
- Provide more explicit differentiation between data collection and analysis processes within the methods section, clarifying each approach and separating sections as necessary for improved coherence.
- Include data collection tools in the annexes to enhance transparency and completeness of the evaluation methodology, ensuring comprehensive coverage of methods utilized (and/or explanation of proposed tools NOT used).
- Enhance transparency regarding limitations by specifying their potential effects on evaluation findings and analysis.
- Strengthen the presentation of findings by itemizing or numbering them to improve readability and facilitate clear links to conclusions and recommendations.
- Incorporate a standalone section for lessons learned to comprehensively capture and detail key insights derived from the evaluation.
- Increase focus on gender and equity considerations throughout the evaluation.
- Improve the presentation of recommendations by breaking narrative text into bulleted action points for easier comprehension and implementation.
- Provide complete information on recommendations, including prioritization, responsible entities, and time horizons, to facilitate actionable implementation.
- Include additional details in annexes, such as data collection tools and initiative implementation modalities, to enrich the report's comprehensiveness and utility.
- Address inconsistencies in spelling, formatting, and numbering to enhance the overall quality and professionalism of the report.

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	While the Executive Summary contains all relevant information to inform decision makers, it is overly long at just over seven pages. Much of it could be truncated, for example the intro/context, the methods/design, the findings (which are in fact the conclusions) could be condensed into bulleted points and the recommendations could be shortened to a single page (narrative guidance summarised in a few bulleted action points).
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	Most of the specified elements are present save the conclusions - which are present, but labelled as findings - and there are no lessons learned. This reflects the absence of a specific lessons learned section in the main report, which is itself contrary to the objectives of the evaluation (discussed below)
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	Insofar as the summary stands, it does present adequate information needed to understand the partnership and the evaluation, but it is not well-written. It is too long, has many formatting and spelling errors and appears to be simply a copy-paste of salient elements from the main report, rather than a genuine attempt to summarise the main report.
SECTION B: BACKGROUND (weight 5%)		50%	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	Section 2.3 provides a good overview of the Blueprint partnership itself, with a good summary of the overall vision and purpose, the itemised outcomes, the locations, timeline and a theory of change. The concluding sentence of the section references the Results Framework for the initiative, but does not include it (or make reference to where it may be found) - this may be deliberate or an omission, but it is suggested to included it as an annex to the report. Section 2.4 provides more granular detail on the specific activities of the partnership. This is useful, but incomplete, insofar as it provides an itemised list of global level joint activities, but only summarises country level activities in a narrative form. Some more specifics on country level would be useful (bullets, the same as for global level, would be fine) - even as an annex. The implementation status is clear (as the initiative under evaluation concluded in 2022) although there is reference to "Blueprint 2.0" in 2.4.4 (pg. 16) which is not clear - is this a second phase? If so, when did it start? The narrative in 2.4.4 appears to discuss implementation of the original phase, with no dates for the second phase noted. Finally, there is no information regarding the budgeting of the initiative. It is not fully clear regarding the "how" of the activities.
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).	Partially	There is relatively little description of the situation of the rights-holders which the initiative purports to assist, and no real description of duty-bearers beyond UNICEF and UNHCR (and that at a relatively high level). The introduction notes the cumulative refugee burden of the participating countries and the cumulative number of refugee children expected to be assisted, but no more granular description is provided. While the evaluation covers an interagency organisational initiative, detailed exploration of the intended rightsholders is not warranted, but, as noted below, some more detail regarding the implementation aspects of the programme (covered under Section 2.3) and the inclusion of the results framework (noted, but not included) would help to ground the evaluation report in some more concrete outcomes that would benefit refugee children.
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 context is reasonably well described - the introductory section (1.1) and the more detailed explanation of the overall institutional and policy context to Blueprint (covered in Section 2) provide a comprehensive and useful background that fully explains Blueprint and the context of the evaluation approach. There is a repeated passage between sections 1 and 2 (para 17 on pg. 9 and para 18 on pg. 14) which should be removed/summarised in one of these sections.
	ii Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Partially	The context section (2.1.1 and 2.1.2) provides some explanation of "Partnership as a central pillar for the delivery of the GCR and the SDGs" in the discussion of the policy framework which underpins the Blueprint, and indeed the analytical framework which governs the structure of the analysis is derived from an SDG partnership policy guide. However, it does not go into any further depth regarding the specific SDGs related to refugee children, for example itemising relevant SDGs or targets. Hence, while the linkages to the SDGs are positive, they could be more specific.
	iii Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Partially	As noted in 2ii above, there is relatively little information (and no exploration) of the status and needs of the rights-holders that are expected to ultimately benefit from this initiative. Section 1.1 notes that 10 million refugee children/family members/host community members (i.e. presumably 4 'additional' beneficiaries for each of the 2 million refugee children in the target countries, per section 1.1, para 17) will benefit, but there is no more specific information on their specific status, needs etc. While a detailed exposition of this aspect is not warranted for a relatively high-level evaluation, some more details, perhaps in an annex, would be welcome.
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	Section 2.2 summarises the history of partnership between UNICEF and UNHCR, in a clear and concise manner. This is useful to include (and with welcome brevity). However, as described above, the description is limited to these actors. It is clear that UNICEF and UNHCR are the primary stakeholders in the initiative, but equally, much of the work that these agencies do is mediated through partners (government and non-government), and therefore presumably the activities of the partnership would equally be mediated through them - but this is not described. As noted above, some more explanation of the implementation process beyond that in 2.3 and 2.4 would be welcome.
	ii Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	Insofar as UNICEF and UNHCR are the KEY stakeholders in this partnership, their high-level joint activities are well described, particularly at global level (sections 2.4.2, 2.4.3). There is, however, relatively little granularity here - the specific contributions (if differing in any way) are not elucidated, nor are the financial contributions (if any). Also, as discussed above, the specifics of country-level partnership activities are overly-briefly described in the same section.
SECTION C:	EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	63%	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.	Partially	Section 1.3 covers the purpose of evaluation, which is clearly and concisely defined in terms of informing "a new UNHCR-UNICEF global partnership agreement in 2023". There is some lack of clarity, however, around what the role of the evaluation is or has been vis a vis 'Blueprint 2.0' which is incompletely described in section 2.4.4 - the section seems to suggest that the evaluation has contributed to that phase, but the timing/duration of that phase is not given, and it is not referred to in the purpose. Some greater clarity/explanation is required in these sections.
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).	Partially	While Section 1.3 provides a clear description of what the evaluation seeks to achieve (there are no itemised objectives), it appears to differ from the overall objectives presented in the TORs (annexed to the main report). In particular, the report notes how the evaluation is expected to assess how the partnership has contributed to "demonstrated results" in the focus countries (said results are not specified, but presumably related to outcomes for refugee children). This is not one of the articulated objectives in the TOR. Given that the evaluation has been 'iterative' in nature, it may well be that the objectives have been refined since the Inception (the Inception Report restates the objectives/purpose as per the TOR), but this is not described and should be.
	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	Overall, the description of the scope of the evaluation is sound - it is clearly articulated in Table 1 (although the nature and relevance of 'Blueprint 2.0' could be included here also). Further, the evaluation questions and scope are well suited to the overall purpose and objectives of the evaluation (notwithstanding the abovementioned issue around differences between the final evaluation report and the TOR/IR).
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.	Partially	Section 2.3 provides a good overview of the Blueprint partnership itself, with a good summary of the overall vision and purpose, the itemised outcomes, the locations, timeline and a theory of change, although as noted above, specific country level activities are not described adequately. Further section 2.3 references the Results Framework for the initiative, but does not include it (or make reference to where it may be found) - this may be deliberate or an omission, but it is suggested to include it as an annex to the report.
	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).	Yes	The TOC is presented in Figure 2 (pg. 15). This provides a useful snapshot of the initiative (and indeed, is one of the only references to stakeholders outside UNICEF/UNHCR). Further, the analytical framework used for the evaluation is a useful breakdown of the different dimensions of the partnership (beyond the evaluation matrix itself, which is also annexed).
	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.	Not Rated	The Blueprint theory of change was not explicitly used to guide the evaluation as the evaluation was NOT theory-based. However, the TOC and the partnership analysis framework were used at the inception stages as the basis for the development of the evaluation matrix (noted in the annexes). The TOR does note that the TOC and results framework would be reviewed and validated as part of the inception phase of the evaluation, which is outside the scope of this review. Hence this criterion is not rated. There is some discussion in Section 7.1 on changes to the programme compared to the original TOC, but this is very briefly discussed and not central to the analytical model.
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	63%	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	The evaluation questions (provided in the Evaluation Matrix in Annex 4 only - not in the methods (although itemised in the findings)) are useful and appropriate, with the matrix providing a comprehensive breakdown of the questions into components - subquestions and indicators. The evaluation questions are aligned with 'evaluation criteria' but these are not all the IASC DAC Evaluation Criteria - they include an additional criterion: Predictability, and the DAC criterion of 'efficiency' is articulated as 'cost-efficiency'. While this evaluation has some non-standard features, the divergence of the design approach from accepted best-practice should have been explained and justified.
	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	While the matrix does indeed include a breakdown of sub questions into indicators (with associated tools), many of the 'indicators' are full-fledged questions in their own right. For example, under KEQ2, sub-question 2, one (of four) indicators related to this sub-question is "To what extent has collaboration between the two agencies been translated into programme/advocacy results as outlined in the JAPs and evidenced through monitoring systems (partnership results being the primary focus, sectorial results, a secondary focus)?" This is a complete question that would require unpacking to be measured accurately. Many (although a minority) of the indicators in the matrix are similarly phrased. This is an issue which should have been addressed at inception.
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.	Partially	The report includes a robust and complete methods section although could be improved by more clearly differentiating between what is covered by data <u>collection</u> vs data <u>analysis</u> . For example, section 3.4 lists six proposed methods of data collection. Whereas the first three (doc review, KIIs, case studies) are acceptable approaches, the latter three are outcomes or outputs of the analysis. The narrative explanation of these elements is a combination of other approaches and a description of the type of analysis to be conducted, rather than clarifying each. In particular, Section 3.4.4 (Evaluation Reporting), should not be a subsection under the data collection explanation, but should be a separate section - 3.5: Deliverables. The data collection tools themselves are not included in the annexes, which is an omission. They are included in the Inception Report, although interview questions are in the form of a "master question list" from which individualised question sets for stakeholder types were to be derived, so again, not a complete set of tools.

	<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.</p>	Partially	<p>The data sources are largely sound for the type of evaluation that is undertaken. As described in Section 3.2, the evaluation is mainly focused on the institutional dynamics of the Blueprint partnership, and therefore has a particular focus on stakeholders at institutional level. It has applied a number of useful principles to ensure comprehensive coverage of these stakeholders, such as the use of iterative, participatory approaches (use of consultation and feedback loops), bringing in different levels of stakeholders (global, regional, country levels) and a systematic approach to managing and analysis data (both primary and secondary). Weaknesses in the approach (as described in the report) are the apparent confusion (noted above) between data collection approaches and data analysis elements and the lack of the voices of rightsholders - understandable to an extent given the nature of the evaluation, but nonetheless the real-world efficacy of the Blueprint partnership could have been grounded in some perspectives, however limited, of refugees themselves.</p>
	<p>iii Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).</p>	Partially	<p>Section 3.5 explains the sampling process for the evaluation. While it is mostly sound, some of the details provided in this section are vague - the evaluators note that four countries were originally envisaged to be included, but two were ultimately "jointly chosen" - but with no explanation of why only two, and what criteria were applied to select them. The section notes "purposive sampling" of countries, but does not explain this. Further, the section discusses including of countries in Blueprint as a partnership - NOT inclusion in the evaluation itself - this is unclear and confusing. Clarification should be provided.</p>
	<p>iv Clear and complete description of the methods of analysis.</p>	Yes	<p>The report gives a very comprehensive account of the methods of analysis of data from different sources - sections 3.3, 3.4 and 3.7 (and particularly 3.7.3) describe the various ways whereby primary and secondary data was treated and analysed for the evaluation itself. Section 3.7.3 specifically addresses the issue of triangulation, breaking it down across data types, data sources, tools applied and the participation of (albeit unspecified) stakeholders to validate and amend findings/analysis. This is useful and clear.</p>
	<p>v Methodology allows for drawing causal connections between outputs and expected outcomes.</p>	Yes	<p>The strength of the planned analysis (if executed as stated) and the relatively comprehensive sampling of stakeholders should lead to a clear articulation of causal connections between the different output/outcome elements under evaluation - this indeed is central to the evaluation purpose, in order to optimally inform the future design.</p>
	<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).</p>	Partially	<p>Limitations and constraints are framed in the context of the <i>evaluability</i> of the initiative, which Section 3.6 notes was undertaken in a 'light-touch' manner. This exercise was undertaken at inception, with the outputs from it summarised and presented in the evaluation report. Explanation of how some of these limitations materialised is given in the report (Table 4), which is positive, but not all of them were explained thus. Further, the table would have benefited from some indication of the extent to which the limitation impacted the analysis/findings. For example, the first limitation notes the challenges involved with progressively limited stakeholder engagement across data collection rounds - but would have benefited from noting, even briefly, how/the extent to which this might have affected the evaluation findings.</p>
<p>Question 10.</p>	<p>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:</p>		
	<p>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).</p>	Yes	<p>The evaluation required and implemented a minimal approach to ethics as it did not include rights-holders. The measures described (essentially anonymisation of respondents) meets the minimum standard. The United Nations Evaluation Group's code of conduct for evaluations was noted in the relevant section. Further details on the different standards that were conformed to are provided in the annexes. This is appropriate.</p>
	<p>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.</p>	Yes	<p>Section 3.8 discusses the ethical approach to the evaluation. This is good, but brief. This said, the nature of the evaluation (focusing on institutional stakeholders) means that the standard of ethical safeguards was minimal, and the evaluation did note the use of UNEG guidelines and appropriate anonymisation of data/confidentiality assurances among respondents.</p>
<p>Question 11.</p>	<p>Does the evaluation incorporate innovative practice that adds value to the evaluation process?</p>		

	i Innovation practice is used to improve the quality of evaluation process. This could 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	While no examples of specifically innovative practices are articulated in the evaluation methods, the Inception Report does make note of two practices that can be deemed innovative: 1. The use of "Stories of Change" (a variation on outcome harvesting or most significant change) to gather qualitative data on changes of importance due to the partnership (described in Annex 6 in the IR). 2. The application of the "Partnership value analysis tool" to analyse partnerships in a combined qualitative/quantitative manner (described in Annex 8 in the IR). Both of these approaches constitute innovative practice, but are, very curiously, not referenced at all in the final report. While they clearly did not form discrete analyses of their own, it is unclear whether they were utilised to support the overall analysis in any way or were ultimately not used. The evaluators should clearly note whether or not they were used - if not, why not, and if so, how.
SECTION E:	EVALUATION FINDINGS (weight 25%)	100%	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.	Yes	This aspect of the report is strong. The layout is clear, with explanatory text before each subsection explaining the elements of the evaluation matrix that are being addressed. This makes for generally easy reading. Thus, the report generally follows the analytical approach laid out in the evaluation questions/evaluation matrix, with headings/subheading following the evaluation matrix themes and bringing in relevant evidence to answer each evaluation question or sub question. The evaluation methodology does note that the analysis has not been fully systematic - it is explained in section 3.3.2 (pg. 21, para 36) that some analytical elements have been combined. This is acceptable insofar as the evaluators explain the rationale for it and it does not affect meeting the overall evaluation objectives. One issue is that the findings themselves are not itemised/numbered. In a report with a complex structure such as this, it would be a useful tool to help readers navigate its complexity, as well as facilitating clearer links between the findings, conclusions and recommendations - the absence of which is a weakness of the report.
	ii Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Yes	As noted above, analysis of the results framework/TOC was not part of the final evaluation scope. There is some discussion in Section 7.1 on changes to the programme compared to the original TOC, but this is very briefly discussed and not central to the analysis.
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	The findings (Section 4) does a reasonable job of presenting the available data in support of the findings. As noted above, the evaluation methodology notes that some elements (i.e. subquestions) have been combined. While simply highlighting and justifying the gaps is not as good a practice as actually providing a comprehensive analysis, the efforts made by the evaluators to address the issue of what could be repetitive findings (to better focus on more fruitful areas of analysis) are laudable. There is good use of data from different sources to highlight analyses. For example, there is considerable use of boxes to provide relevant examples from country context to highlight analyses based on primary data from interviews that are equally well-supported by an analysis of available documentation.
	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	Much of the analysis and findings are supported by good and well-referenced evidence noting if findings came from country offices, for example. This said, the analysis varies in quality in this regard, featuring robust and well-supported reasoning in some instances, but also displaying repetition. Some sections are not fully aligned with the analytical framework questions/sub questions/indicators, despite presenting useful and informed analysis. For example, Section 5.2 (Achievement of partnership results), while interesting, lacks discussion of any specific results reported by the programme, and also lacks a clear link to Section 7 which discusses concrete outputs in more detail (although para 97, pg. 39 does note that Section 5.2 focuses on a more high-level approach to results whereas Section 7 focuses on specific outputs).
	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	In many places the evaluation seeks to determine the causes or determining/inhibiting factors for achievement of results by exploring various dimensions of the data. For example, under 5.4.2, analysis from the document review is presented relating to fundraising modalities, but primary research among countries indicated that such modalities were not successful (including a specific country example). The section uses interview data to analyse the reasons for this, presenting a number of important challenges that should be addressed to operationalise the (unsuccessful) fundraising modalities. This is a well-reasoned and useful piece of analysis.
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.	Yes	The approach taken to monitoring and reporting Blueprint results and the implication for the evaluation is discussed in section 5.2.3. While brief, (approximately 1 page) the section does provide some useful insights, including a table of key challenges and adaptations around monitoring exhibited by the programme. Further, the results, effectiveness or specific issues related to the monitoring system are referenced in other places, such as the monitoring indicator related to cost-efficiency of the initiative (discussed under Section 5.5, para 130). Finally, monitoring forms part of the recommendations of the evaluation, and is as such well covered.
SECTION F:	EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)	83%	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	While quite dense and narrative in nature (the conclusions span four pages of narrative text), they are a sound summary of the analysis and findings and present positive feedback that clearly signposts what has been achieved well and less well, and thus indicates what should change for a future iteration of the UNICEF/UNHCR partnership. While shorter statements or summary single-sentence conclusions before each paragraph may be more readable, the existing formulation is adequate and may be agreed or preferred by UNICEF.
	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	Overall, the conclusions section is very well-written, thoughtful, and draws on many of the findings articulated across the main analysis, although they are not itemised with reference to specific findings (which themselves are not itemised or numbered). It is structured in accordance with the evaluation questions and clearly described as such. As noted above, the conclusions provide a comprehensive picture of both the positive and the negative of the initiative and should be useful and acceptable to both UNICEF and UNHCR.
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.	Yes	The report does not have a standalone good practices/lessons section but has highlighted lessons via bulleted sentences interspersed throughout the text of the report. While the level of effort applied to identification and description of the lessons is questioned (discussed below), those that are presented are indeed derived from the findings and have wider applicability, although there is a balance to be struck between having wider relevance and being overly general.
	ii Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	There are many good practices and/or lessons articulated throughout the report, with 11 of these highlighted in-line in the findings. This is positive, and they are clearly presented. However, although this minimally acceptable, one of the two key objectives of the evaluation was to "Identify lessons learnt and good practice[s]". This would suggest that evaluators have an obligation to make particular efforts to capture and present such lessons - interspersing them amongst the findings is not the optimal approach, particularly given that UNICEF report guidelines mandate a specific lessons-learned section. The preamble to the conclusions section notes that it includes lessons, but they are not highlighted in the narrative. Further, one of the recommendations is that efforts should be made to capture more of the lessons of the initiative, noting that " <i>In addition to the lessons identified in this evaluation, there are others that are specific to each agency. There would be much value in gathering all of these lessons together</i> ". This appears to have been the role of the evaluation, which has been only partially fulfilled. In short, to satisfy the clear objective of the evaluation in this regard, a specific lessons-learned section that provides extensive detail on the many lessons and practices that have clearly been gathered (possibly via the Stories of Change approach cited above) would have been warranted.
SECTION G:	RECOMMENDATIONS (weight 15%)	75%	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.	Partially	The recommendations are a good mix of three different elements of partnerships (fundamental, management and relationships) that are relevant to the evaluation itself. The text under the recommendations is somewhat dense, however - breaking the blocks of text into bulleted key action points would have been a more useful approach to ensure that they are more easily digested into appropriately-implemented activities. While the recommendations do include a reference to the specific section where the analysis that supports them is located, there is no specific link to either conclusions or individual findings - this inevitable given that neither findings nor conclusions are itemised/numbered in the report.

	ii Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Yes	The recommendations are well-structured - with headline statements followed by more detailed explanations, although rather than blocks of narrative, bulleted action points targeted at specific agencies/divisions might have been more useful. They are largely actionable and realistic, with priorities and key actors identified for most, (though for some they are TBD and the time horizons for each are absent).
	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).	Yes	The report describes how the findings (and recommendations) underwent an iterative process of review and revision involving a variety of stakeholders at global/regional/country levels. It therefore appears to offer a robust level of consultation between the different stakeholders involved. Rights-holders were not involved in the evaluation.
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	As noted, the recommendations are split into three broad areas - this is an appropriate strategic division of them. They are also prioritised according to high/medium/low, with the appropriate responsible entity noted (for some - some are still 'TBD'). The allocation of recommendations to responsible entities is incomplete. Also, the time horizon for recommendations has been deleted from the previous draft - this should be reincluded and the missing data included.
SECTION H:	REPORT STRUCTURE AND PRESENTATION (weight 5%)	58%	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.	Partially	Most of the specified elements are present in the opening pages, save the period covered by the evaluation (2020-2022).
	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	There are four annexes - the TORs, key informants, bibliography and evaluation matrix. These are useful, but the report would have benefited from the data collection tools, some additional description of the initiative implementation modalities and the initiative results framework.
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).	Partially	The structure is generally good, with numbered headings/subheadings and paragraphs. However, there is one numbering error with the paragraphs, and formatting issues with headings being placed at the end of pages rather than at the top, and tables breaking across two pages when they should be formatted to be on one. These are easily addressed.
	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.	Partially	As noted, the report omits a standalone Lessons Learned section, which, given the priority clearly ascribed to gathering of these lessons implicit in them being one of the two objectives of the evaluation, was an omission. Otherwise the structure is appropriate.
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.	Yes	The TOR mandates a final report of 40-50 pages (excluding Executive Summary and Annexes). This report is approximately 60 pages, which is acceptable given the nature of the subject matter and the scope of the evaluation and its multiple phases.
	ii Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	The style and writing are generally good, and it is quite readable. There are some style/spelling issues evident in the report. These are notable in the Executive Summary, which is likely to be the most read portion of the document and hence should be presented to a very high standard. Issues such as spelling out of acronyms (e.g. SDGs is explained twice in the glossary of terms on pg. 4) or heading breaks (e.g. pg. 2, pg. 4, pg. 9), inappropriate use of capitalization of words (throughout) and a mix of US and UK spelling formats all detract from the quality of the report. The authors should use a consistent style - preferably following the UN style guide rules. Finally, the paragraph numbering starts over between the end of Section 1.1 and 1.2 (pg. 9).
	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 evaluation report makes good use of a combination of tables, figures and boxes. These are mostly useful and well-formatted, but there are some editing issues. For example, box 2 (pg. 26) is not really needed as a box as it comprises a single sentence. Suggest to just leave it as narrative. Also, some figures are of poor resolution - for example the TOC (figure 1) is text dense, but very difficult to read due to the resolution issues. Finally, some tables break across pages unnecessarily.
SECTION I:	EVALUATION PRINCIPLES (weight 10%)	50%	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 does clearly reference the CRC in the description of the evaluation context (2.1.1, pg. 13), albeit once only, in the description of the policy and normative context in which the Blueprint initiative lies, i.e. that it underpins engagement with the GCR and the partnership with UNHCR. This and the following section clearly link the partnership to such policy benchmarks and frameworks (including the SDGs), which is appropriate.
	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).	Yes	The evaluation has solicited the participation of a variety of stakeholders in both the data collection and in the formulation of findings, conclusions and recommendations. While the use of boxes to highlight findings using the examples of specific countries is positive, the same could be done for the voices of respondents or informants, via quotations from respondents (anonymised) to support the analysis under specific findings. For example, the analysis of partnership relationships in section 8.1.3 (pg. 65) notes a significant "number of examples given during the evaluation when a loss of trust undermined an existing partnership" - provision of some of these examples (via quotes) would have provided useful colour and context to the evaluation, but is not essential. The evaluation did not go to the level of rights-holders
	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).	Yes	The nature of the methodology - primary data from institutional stakeholders only - does not lend itself to disaggregation of results in this manner, hence this category has not been rated. Similarly, there is no discussion of rights-holders or duty-bearers, but there is no indication that the language is in any way disempowering or biased - there are some references to refugee inclusivity which signal appropriateness of language.
	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.	No	Despite the design of the evaluation clearly noting an emphasis on "a gender sensitive approach", ensuring a focus on "gender, diversity, vulnerability and inclusion" and that "Where relevant and possible, [data] [sic] was disaggregated according to these parameters", (all noted in Section 3.4.3) there is little analysis presented in the report around this, either in terms of how UNICEF or UNHCR are mainstreaming these issues in the partnership, or how the initiative has led to better outcomes for rights-holders. Issues of equity, discussed in section 6.2, are more focused on internal organisational dynamics. Section 7 discusses issues of Inclusion - specifically in relation to refugees, which is positive. However, the section is largely focused on examples of how specific initiative outputs manifested in different countries. While there are examples of child-rights, disability or gender-related programming that have been facilitated by the initiative, there is no serious analysis of this and it comes across as more of an afterthought to the report.
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)	3	
	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.	Partially integrated	As noted above, while the report highlights GEEW considerations briefly in section 3.4.3 in the methodology, there is relatively little further explanation or mention in the scope.
	ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Partially integrated	The evaluation methods make note of gender and equity considerations in several places, notably a commitment to "scrutiny of gender and equity aspects of the partnership" as a "data collection" technique (although this is actually data analysis or an outcome of the analysis). However, this is not unpacked further, despite section 3.4.3 noting further commitments by the evaluation. In line with the limited discussion of GEEW in the methods, the tools were not included with the report but are (in summary fashion) in the Inception Report, and have no GEEW elements. Further, the evaluation matrix does not include any mention of gender OR equity.
	iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Not at all integrated	The analysis does not demonstrate any significant GEEW analysis. This is unsurprising given that the, already thin, commitment to GEEW analysis presented in the methods did not make it to the evaluation matrix. In fact, the findings mention gender twice, both in the context of examples of programming from different countries, and not in any deliberate process of GEEW analysis or presentation.

SWAP Rating Guidance

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

- 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?
- Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?
- Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?
- Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?

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

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

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?