GEROS Evaluation Quality Assurance Tool

Version: September 2021

Formative Evaluation of End Child Marriage (ECM) Flagship Results, Ethiopia

REPORT RATING SUMMARY			
Overall Rating		71%	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	Formative Evaluation of End Child Marriage (ECM) Flagship Results, Ethiopia
Report sequence number	Ethiopia/63/2023/18899
Region	ESAR
Year of report	2023
Office	Ethiopia
Coverage (countries)	Ethiopia
ToRs present	Yes
Date of review (dd/mmm/yyyy)	February 1, 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	Programme
Evaluation type	Formative
Evaluation strategy	Mixed methods
Evaluation design *(primary method used)*	Theory-based
Evaluation level	Output & Outcome
Geographic scope	National
Primary SDG(s) covered *(number)*	5.3

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 satisfactory report that may be used by decision makers pending revisions. The complexity of the multi-sector scope of the evaluation is well noted and the findings useful for future programming.

The report offers the following strengths:

•The executive summary is clearly presented within the recommended page limit.
•The context of the intervention includes the major factors leading to child marriages.
•The context notes the linkages to the SDG 5.3 and key national strategies.
•The intervention outputs and outcomes are mentioned and correlate with the results logframe.
•Stakeholders are mentioned receiving technical and financial support through the programme.
•The purposes and objectives of the evaluation are clearly noted as well as temporal and geographic scope.
•The scope reflects the inclusion of gender, equity and human rights.
•Both the Theory of Change and results frameworks are included in the annexes, design and findings.
•The evaluation questions are appropriate for meeting the objectives and purpose and are effectively presented on the evaluation matrix.
•The methodology is mainly strong to test a theory-based approach and the limitations clearly mentioned with mitigation measures.
•The gender and equity focus and ethical protocols are very effectively covered in Annex I.
•The evaluation used a relatively new tool the Basic Efficiency Resource method to measure cost per participant and cost efficiency.
•The findings systematically address the questions and use credible evidence and triangulation, and in most cases strongly present the causal factors.
•The findings present a well developed assessment of the monitoring systems.
•The conclusions are effectively presented by criterion.
•The recommendations are useful and have been well vetted with duty bearers and rights holders.
•The report structure includes the relevant information and is clearly written.
•The UNSWAP criteria are fully or satisfactorily integrated in the approach, findings and recommendations.

The top line issues are as follows.

•The executive summary requires strengthening to be concise, focused and useful to decision makers with findings/conclusions and recommendations strongly linked.
•There is a limited description of the characteristics of the target populations including the humanitarian assistance situations of some rights holders. (Section 1)
•A general description of the UNICEF sections and their roles in supporting girls' development as a basis for the ECM Flagship is missing.
•Budgetary figures are missing as is a mapping of the stakeholders, their roles in the programme and the evaluation and the linkages between them.

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 recommendations for improvement are the following.
1.Expand the description of the characteristics of the target populations including the humanitarian assistance situations of some stakeholders. (Section 1)
2.Include a general description of the UNICEF sections and their roles in supporting girls' development as a basis for the ECM Flagship, as found in the TOR, in an object description which should form a separate section.
3.Include a clear approximation of the budget in a separate object section as well as a mapping or matrix of stakeholder, their roles in the programme and the evaluation and the linkages between them.
4.Clarify the purpose as a mid-term exercise with end date of the programme 2025 and include the programme budget. (Section 2).
5.Clarify the Theory of Change as a draft that was not revised by UNICEF or the team and the evaluation use of the logframe outcome and outputs. (Section 2)
6.Describe the process to agree on the evaluation questions and any rewording or regrouping of the TOR questions (Section 3)
7.Justify the acceptance of the UNICEF recommended sampling of the regions as described in the TOR.
8.In the Findings (Section 6) provide the disaggregated targeted numbers for Figures 6, 7, 8 and 9, such as whether in humanitarian settings.
9.Clarify that the analysis presented in the Efficiency findings as a cost effectiveness analysis is a cost efficiency or cost per participant analysis. (Section 6)
10.Revise the conclusions section (Section 7) to separate out the findings, conclusive statements and lessons and reduce text to succinct highly analytical conclusive statements of strengths and problems that will launch the recommendations.
11.Create a separate section for the lessons and express them as key generic lessons in bolded first sentences followed by experiences from the programme that support the lesson.
12.Review the recommendations (Section 8) to firmly link to the conclusions and include clear and well thought out steps for implementation.
13.Include the missing annexes: data collection instruments, list of interviewees, names of woredas and sites visited, and move the document list to the annexes.
14.Strengthen the executive summary to highlight the objectives, findings/conclusions and recommendations. Reduce background discussion to a few sentences, summarize the methodology and mention the sampled regions and numbers of interviews. Strengthen the conclusive statements ensuring they link with the recommendations, and note the recommendation priorities. Revise as needed based on the recommendations from this review in Sections F and G, and include a few key lessons in the summary.

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 summary is clearly presented within the five page limit. The background discussion is not concise and cut and pasted from the background in Section 1 taking up an entire page. For the purpose of informing decision makers regarding the evaluation process and results, the main focus should be on the findings, conclusions, recommendations and lessons. A better balance should be achieved, with the background offered within a few sentences which is actually not required. The discussion should move quickly on to a tight summary of the intervention, particularly the theory of change in regard to intended results and outcomes, all within one or two paragraphs. (Note: the sources are not needed for exec since they are in the main text and disrupt the flow needed for executive reading.) Suggest to consider placing a matrix, revised according to the recommendations below that clearly indicates the recommendations, their priorities and the responsible parties.
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 summary includes the overview, the objectives, and intended audience, the methodology and key conclusions and recommendations. The purpose should be distinct from the objectives and suggest to include the reason for the evaluation at this midterm point in time. The objectives should be offered in more detail as the decision makers will look at the conclusions and recommendations through the lens of what the evaluation was meant to assess. Suggest to include the objectives that relate to assessment of synergistic effects and partnerships. The objectives may also be bulleted as found in the TOR. Once the lessons have been extricated from the conclusions as recommended in this review, several could be added, as lessons are requested in the TOR.
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 summary does not introduce new information however, it is noted that more conclusive information on the cooperation of the specifically mentioned sectors is found relative to the conclusions sections in the main report. More detail is needed to understand the intervention, its coverage and numbers served, its intended results, the multi-sector effort, expanding the text, while the methodology can be tightly summarized in one paragraph. However, it is important to mention the regions that were sampled and the numbers of interviews conducted. There is no need to mention 'findings show" or "we found" and to avoid awkward wording which diminishes the clout. Suggest to examine the paragraphs to ensure that they are tightly expressed, and are similar to the revised conclusions as recommended below in this review. Convergence is discussed in relevance, efficiency and effectiveness and reasons should be clear relative to the criterion, such as for relevance, "limited multi-sectoral convergence reduced relevance of the combined interventions in supporting girls' developmental needs". For efficiency, the outputs achieved could be mentioned. To introduce the recommendations, suggest to add one sentence indicating the process to agree on them with rights holders and duty bearers and note they are mentioned in order of priority. Ensure that the recommendations are clearly launched from the findings/conclusions and mention the problem or strength, such as specific weaknesses in monitoring capacity and data collection, the strengths or effectiveness of peer to peer and gender clubs for mentoring and peace building so the recommendations are well grounded in findings and not merely good ideas.
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	The description of the object begins on page 11 after description of the context in the background section (Section 1). The End Child Marriage (ECM) is a global program and in Ethiopia partnered with UNFPA and implemented through government and Non-government partners in the (UNICEF) ECM Flagship Result Programme. The current phase focusing on 2020-2023 is the phase under assessment and covers all eight regions including some humanitarian settings. The programme budget is unclear. Suggest that the details of the programme are placed in a section titled Subject or Object of the Evaluation (as found in section 2 of the TOR) to draw attention to the programme itself as normally part of UNICEF report structures. Include the budget for the programme, and the geographic scope. The humanitarian aspect seems to be intersected and this should be explained both in the context and the Object as humanitarian populations are featured in the results indicators. What is missing from the description are the activities of each of the sectors and this is important to understand what the evaluation will assess. Include the list from the TOR page 77 that is more summarized as to the specific actions that each section is expected to implement.
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	Page 12 includes a paragraph noting the institutions and community members who receive technical and financial support but are not referred to as duty bearers. Rights holders are only briefly mentioned, such as out of school girls. The report jumps to the methodology without a clear description of the stakeholders in the object in terms of duty bearers and rights holders. The numbers of rights holders (targeted) as presented on page 76 of the TOR are not brought into this report. These include a humanitarian assistance targeted population. The picture is thus not specified or clear as to who the rights holder are, how many, disaggregated as to age, disability, as part of humanitarian assistance targets, among others. This would be expected to be presented on a matrix or narrative fashion with due attention to and is not found in the inception report either. Please bring in the numbers of targeted people and their characteristics to add to the description of the rights holders and 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.	Partially	Section 1 describes the context which includes the situation of child marriage in Ethiopia, the generalized harmful effects of child marriage and the factors impacting child marriage in Ethiopia with linked discussion of the UNICEF and UNFPA programme. It would be useful in the context to more fully describe the extent of the problem specific to regions such as which locations and populations have a greater number of child marriages and why, such as noting the conflict and drought affected areas, noting the most vulnerable age groups, among others. This would link in to the sampling strategy discussed later (Note: it is more effective to use footnotes with complete sources rather than insert in the text which adversely affects readability.)
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Partially	Linkages are drawn to the SDG 5.3 and two paragraphs are devoted to the main national strategy, the costed roadmap to end ECM and genital mutilation (date missing) and ECM Flagship contributions to UNICEF strategic objectives. It mentions Article 35 of the Constitution (date missing) which guarantees women the right to protection from harmful traditional practices. A national law is mentioned in paragraph 1 which may be Proclamation No. 414/2004 mentioned a few paragraphs down but this is unclear. It also refers to Ethiopian government policies 2005-2016 which have presumably been effective in reducing the prevalence and incidence, but unclear which ones these are. (Note: Error in mentioning that Ethiopia came to power in early 1990's and should be clarified, the current government, the constitution, or other.) Please clarify the name of the national law and some details such as punitive measures and how well they are enforced, which link to the efforts of the programme, as well as and human rights instruments that are relevant and the accession of the Ethiopian government to these instruments. These would include the relevant targets for gender equity, child rights, poverty alleviation, humanitarian assistance providing a broader perspective on how the programme supports these goals. Since the programme is multi-sectoral, what are the conditions, the standards and the issues in protection, WASH, education, health, nutrition, emergency, social protection and communication? Suggest the context section includes a general description of these sectors as they influence the targeted rights holders toward or away from child marriage and then expand the discussion of the

iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Partially	There is a general description of situation of rights holders and the numerous factors that influence child marriage. However, there is weak indication of the understanding of how the programme outputs are going to meet the needs of the rights holders except in a very general way. The Terms of Reference offers a stronger picture of rights holders in Ethiopia, such as the numbers of persons with disabilities, This would contribute to understanding how the programme targets the needs such as why it targets girls 10-19 years of age, why they need multi-sectoral services, and why the duty bearers need capacity development, such as to enforce the law and follow the roadmap. Further there are gaps in identifying the characteristics of the targeted girls, such as how many in school and out of school, how many never married, how many divorced or separated, practicing prostitution, how many affected by emergencies and may be living in humanitarian situations such as temporary settlements, those whose families receive humanitarian assistance, sources of income for their families such as pastoralism, farming and other characteristics that would have a bearing as to the family and individual pressures to engage in early marriage. Taking for example the first output of the programme "underserved marginalized adolescent girls" please explain who these are and details about why or how they are underserved and marginalized
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	There is mention of partners such as UNFPA and government ministries. Given the numerous stakeholders and their complex linkages, there is no mapping or organization on a matrix of the key stakeholders and this is not found in the inception report either. Suggest to refer to other UNICEF reports where stakeholder matrices or mappings have been developed with columns for the category of stakeholders, their roles whether financial, oversight, monitoring, capacity building, partnerships such as managerial or implementing or UN or other similar, their interests in the evaluation, and their linkages to each other.
ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	A paragraph appears on page 12 and some information in others regarding UNICEF's role and the implementation arrangements and co-funding by the Ethiopian government and its relevant ministries and regional and woreda level bureaus. A more organized and developed description is needed and generally a stakeholder matrix presents this clearly with the important information for readers on the contributions and roles and linkages to each other. The role of UNFPA is missing and this is unacceptable as it is a key partner. (Note: HTP is not defined and does not appear on the acronym list.)
SECTION C:	**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	75%	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 2 describes the purpose and objectives and scope. The purpose stated is to draw lessons, assess and document the relevance, effectiveness and efficiency of the programme to inform future efforts, contribute to strategies and multi-sectoral cooperation throughout and beyond the implementation period. It is important to note why the evaluation is needed at this point in time. Suggest to include from the TOR, Section 3: "With the CPD approaching the midterm - amidst Covid-19, conflict and droughts - this evaluation is timely in looking at what has been done for Ending Child Marriage (including challenges faced and how the programme has mitigated), and in looking forward at how ECM flagship can accelerate the progress towards ECM." The evaluation is taking place in the middle of the programme as it ends in 2025, this is important to understand relative to the purpose. The key intended users stated are UNICEF and implementing partners. Since the key partners will read the report, it is important to note them specifically, suggest to include from the TOR: "Evaluation's expected users are mainly UNICEF Ethiopia Country and Field Offices, UNFPA, government counterparts (including Ministry of Women and Social Affairs (MoWSA), and Ministry of Education (MoE)), other implementing partners and key stakeholders. Secondary users are actors in the wider community working on ending child marriage."
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	There are five objectives and are the same as in the TOR, which should be mentioned. These objectives are very important and nuanced as to what exactly is wanted, and along with the purpose they need to be reflected in the executive summary to link to what the process has assessed in the findings, conclusions and recommendations and the lessons learned.
ii	Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	The scope is clear and limited to the programme and all its aspects including the multi-sectoral contributions which is a complex scope with at least nine UNICEF sections involved. Importantly, the gender, equity and child rights dimensions are considered. Temporally, the scope is July 2020 to February 2023. Data was collected in the desk review from all eight regions but three were sampled Oromia, Amhara and Somali as suggested by UNICEF.
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	This is a theory based evaluation, Annex VI shows a Theory of Change diagram which is slightly blurry, and Annex VII presents the ECM Flagship Logframe. Page 13 presents the expected outputs of the ECM Flagship. It appears the evaluation will assess progress on each of these, but that is not completely clear. A footnote indicates that there are inconsistencies between the ToC and the logframe with the ToC depicting more outcomes and outputs than the logframe and the team has selected to use the logframe one outcome and four outputs. Suggest to include an explanation as to how the results-based tools affect the findings as the findings are expected to refer back to the ToC for theory based evaluations. An explanation is not found in the inception 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 report presents the logframe in Annex VII and the Theory of Change in Annex VI. The logframe outcome and outputs are used in the evaluation matrix under the evaluation criteria with indicators.
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	The evaluation team has assessed the ToC and the logframe and opted to use the logframe with its one outcome and four outputs which is also used in the TOR. The TOR notes that the Theory of Change is in draft which is not mentioned in the report. It is unclear whether UNICEF or the evaluation team might have revised the ToC to match the logframe and finalized it as might have been requested by the evaluation management. These are different tools with the logframe being part of the overall country programme, whereas the ECM Flagship ToC may be a more accurate and focused set of impacts and outcomes. Please add an explanation regarding the use of the draft ToC and why it was not revised by UNICEF or the team.
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? *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	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 are discussed in section 3.1. The Coherence criterion and question proposed in the TOR is not present. "To what extent have the outputs of the ECM flagship result been achieved or are likely to be achieved efficiently through the linkages and synergies between sectors?" The report should mention the process undertaken to agree on the questions with evaluation management and the evaluation reference group and note where changes have been made from the TOR. Since an objective of the evaluation is noted as assessing synergies and partnerships (Objective #1) the regrouping of this question requires explanation, such as whether it became cross cutting.
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	Annex III features the evaluation matrix. It includes the questions, the indicators, the sources and whether the data is qualitative or quantitative. Typically the matrix notes more specifically what types of analysis will be conducted from the methods described and how sources will be triangulated. Suggest that the last column to the right is completed more thoroughly to show how conclusions will be drawn or another column added.
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 discussion on data collection methods is separated into primary and secondary, which is useful. Section 3.2 covers primary data collection and section 3.3 covers secondary data. The data collection tools for the primary data are presented in the evaluation matrix and include questions and indicators tailored to the respondent groups. These tools are not found in the annexes and should be attached to the report. They should be referenced in section 4 on data collection.

ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Partially	From primary data sources, these include the rights holders and duty bearers including UNICEF staff to some degree. It is unclear why UNFPA would not be included or other UN or NGOS working for ending child marriage as this would likely enrich the findings. This also points to the lack of a stakeholder matrix which would explain the interlinkages. The data sources are typically not complete without the external stakeholder inputs. They would necessarily include the evaluation users. Suggest to find a way to include UNFPA if not already and key organizations such as Save the Children working for the same goals and interweave their opinions.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Yes	The evaluation design and methodology section (3) includes a section 3.2.1 on sampling strategy and data collection tools. The three out of the eight implementation regions where interviews were conducted were primarily selected by UNICEF. The team then used purposive sampling to select two woredas per region. There are five selection criteria for the woredas and the woredas are mapped on Figure 1, which is very useful for readers. The sampled woredas population was further stratified into respondent categories which included female adolescents. There is a very strong rationale set forth by UNICEF for selecting the three regions and it is important to mention these and that the team agreed with the rationale as usually the team is free to sample without bias or influence. This rationale is well explained in the TOR, page 79 and suggest it is brought in verbatim. Table 3 in Annex II provides an overview of the persons interviewed.
iv	Clear and complete description of the methods of analysis.	Partially	Section 3.2.2 discusses the quantitative Analysis which is a two-stage process, first the team's reflection and then a more thorough process including coding data and looking for themes is mentioned. The second part is not fully described such as whether certain programmes for coding and data sorting were used for the qualitative data. Please be more specific in describing the analysis procedure for the qualitative data, such as how the coding was done and if possible include a summary of the products in the annexes. The process for analyzing the secondary data is more thorough, with quantitative being processed through Stata. The Value for Money approach and using the Basic Efficiency Resource method was employed for cost efficiency, as relevant expenditure and outcome information was missing (Note: should be listed as a limitation.) Data triangulation is included in both primary and secondary and cross cutting.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Partially	Since it is a theory based evaluation, the methods of analysis descriptions should be clear on how the inputs to outputs and outcomes will be analysed and will note where the ToC and logframe have represented the causal connections accurately or they require revision. Suggest to include more description of how the team traced the causal connections.
vi	Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Partially	Section 5 covers limitations and research gaps. The constraints faced are well described as well as the mitigation measures employed by the evaluators. As per the discussion in 3.3.2, the issues with absence of relevant information on expenditures and outcomes should be part of the limitations discussion.
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).	Yes	Annex I covers Gender Equality and Empowerment of Women and Ethical Considerations and presents a very comprehensive picture of respect for guidelines and standards. It describes respect for UNEG guidelines and Ethical Research Involving Children and includes descriptions of the ethical principles that the team considered in the evaluation design. Gender and human rights have been incorporated throughout. The integration of gender and empowerment of young women is described.
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.	Yes	Annex I describes the submission of forms and consents to the independent ethical review board and receiving the ethical clearance in January 2023. Further the consultancy group C4ED's principles are described on privacy and confidentiality. The last paragraph of section 3.2 on qualitative data collection mentions upholding principles and standards in the development of the data collection tools and Annex 9 presents consent forms for respondents including children. Typically the letter(s) of clearance are attached to the report. (If possible please clarify which authority granted ethical clearance and attach the letter(s).)
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 evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	No	The use of the Value for Money approach and the Basic Efficiency Resource method is a recently publicized tool particularly when data is missing for assessing cost effectiveness. It is to be noted however, that the evaluation does not compare the intervention to other similar interventions and their costs and these tools may be acceptable for cost efficiency analyses comparing the programme activities to learn about cost per unit or stakeholder perception of good or wasteful use of funds and other resources. (Note: the technical definitions of cost effectiveness and cost efficiency analyses differ with cost effectiveness comparing costs of two or more courses of action; and cost efficiency looks at delivering inputs at the lowest price without compromising quality which is what this evaluation seems to have done.)
SECTION E:	**EVALUATION FINDINGS (weight 25%)**	83%	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	Section 6 presents the findings which are organized by criterion with the evaluation questions mentioned and answered in order. The findings marshal the evidence and systematically answer the questions. There is reference to background information, at times to excess. To answer Relevance EQ 1.1, "To what extent is the ECM Flagship Result Programme design and implementation relevant in addressing major issues for girls' vulnerability to Child Marriage", the findings present Figures 2-5 which seem to be background statistics regarding the incidence and prevalence of child marriage and unclear whether the team did more research over and above the information presented in the background section. If this research was taken from secondary sources, the sources are not mentioned. Please footnote the sources of the information/findings presented and whether from secondary or primary sources. The alignment of the priorities discussion presents four paragraphs of factual background (pages 26- 27) on the Ethiopian government policies which may be relegated to the background section whereas the actual findings do not appear until the last paragraph on page 27.
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Yes	The findings section includes reference to the ToC in the efficiency section and notes weaknesses in the monitoring system due to inconsistences between the ToC and the logframe. The ToC is examined in terms of effectiveness in immediate and intermediate outcomes.
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 are presented using both qualitative and quantitative data. For example, the Efficiency section in discussing the contributions and synergies refers to the ToC and its assumptions as well as the logframe Output 3 and its indicators, however, the Efficiency section has definitional issues as described in Question 13.ii below. The sources are triangulated.
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	The findings are supported by the evidence and both positive and negative findings are presented and indicators, standards and benchmarks are drawn in. In the Efficiency section, there is evidence provided on the cost per person of each of the output indicators. The section explains why the analysis suffers limitations. Technically, this is not a cost-effectiveness analysis as it is labeled and thus should be re-labelled as a cost benefit or cost per output analysis. (Note: the technical definitions of cost effectiveness and cost efficiency analyses differ with cost effectiveness comparing costs of two or more courses of action; and cost efficiency looking at delivering inputs at the lowest price without compromising quality which is what this evaluation seems to aimed for.) The qualitative findings presented on cost efficiency are more valuable as to the resource constraints identified by the respondents.
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	The findings clearly discuss the causal factors and also analyze the logical chain. For example, the Effectiveness section assesses progress on achieving the performance indicators. The section also identifies unintended effects which included increased community engagements, shared savings though gender clubs, and some instances of more assertive women's roles.
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 findings section presents a well developed discussion on the monitoring system, both the country programme and the ECM Flagship, pages 34-37. Inconsistencies between the ToC and logframe and its effects on monitoring are discussed. The ECM M&E system is described as a robust framework which met with constraints and weaknesses in its implementation such as weak monitoring capacity, and lack of baseline data and disaggregated data.
SECTION F:	**EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)**	50%	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).	Partially	The conclusions (Section 7are forward looking and valuable. However, some read more as a summary of findings and already include recommendations or suggestions. Suggest to introduce the conclusions section by repeating the evaluation purpose and objectives in order to allow readers to link them. Suggest to remove language such as "we found" or "findings indicate' as these are not needed in a conclusions section, and transform the statements to be highly conclusive, succinctly pointing out the strengths, gaps or weaknesses that will launch the recommendations. In the first bullet point under Relevance, the first two sentences are conclusive, however, a tentative recommendation follows. It would be more effective to lead into a recommendation to mention why UNICEF has not partnered or that partnerships are missing that could lead to more effective results. The third bullet point contains largely recommendations rather than succinct problem statements that will lead to the recommendations, thus lacking the clout needed. The first objective of the evaluation: "To assess the Flagship Result Programme design and results contributions of various actors, as well as the synergistic effects of the various interventions (from several sectors) to ECM (at outcome level)" seems to not be clearly addressed given the various contributions from the different sectors, such as WASH, Protection, Nutrition, Education, among others.
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.	Partially	The conclusions are derived from findings and present strengths and limitations. However, the conclusions section contains excessive discussion of findings, rather, the strengths and problems should be expressed with highly analytical (i.e., conclusive) language. (Note: Summaries of the findings can be very useful and placed in the findings section.) The Efficiency first paragraph does a good job of consolidating the conclusions. The bullet points then mix lessons with findings. For example, the fifth bullet point presents a number of wandering points and suggestions, whereas the main conclusion is buried. This may be: "Significant field capacity is absorbed by emergency response, which limits the effectiveness of the ECM Flagship interventions including insufficient human resources for strong engagement and efficient monitoring resulting in weak linkages to programme participants." This statement or similar is adequate to launch a recommendation. However the listing of sectoral contributions to the ECM indicates that emergency response is to include the ECM interventions and this should be alluded to.
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	Lessons as presented can be valuable but they are mixed in with conclusions and are generally not in the correct form for lessons. This burdens the conclusion section and serves to inadequately highlight both the conclusions and the lessons. Create a separate section to present the lessons and ensure that the key statement is expressed generically to be relevant and applicable to other programmes.

ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	The Conclusions section presents bullet points under the lead criterion paragraphs which are termed lessons but some are conclusions and some may be lessons while others are findings. Suggest sorting these out to extract the conclusions that are going to lead to the recommendations and ensure they are in the Conclusions section. Select those that should serve as lessons and can be expressed in generic terms with some supporting discussion from what was learned in the implementation. Under Effectiveness, the first bullet point is written as a conclusion while the second is expressed as a lesson. Under Sustainability the first bullet point offers a lesson statement which is then mixed in with a conclusion. The third bullet point mixes findings with lessons and conclusions. For example, this statement can be transformed into a lesson by adding a few sentences on what was learned through the ECM experience: "Promoting gender equality and strengthening community mobilisation and response mechanisms for ECM can contribute to increased community engagement and more equitable distribution of roles and responsibilities beyond the area of child marriage." If it is to be a conclusion, it can be expressed differently, for example: "The promotion of gender equality through ECM Flagship has contributed to community engagement to address CM and as an unexpected benefit has contributed in some cases to more equitable distribution of roles and responsibilities among males and females."
SECTION G:	**RECOMMENDATIONS (weight 15%)**	75%	:
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 (Section 8) align with the evaluation purpose to "inform future efforts and contribute to strategies and multi-sectoral cooperation throughout and beyond the implementation period". As such the first recommendation addresses multi-sectoral convergency which springs off the Efficiency conclusive statements. However, to firmly link the conclusion to the recommendation, it is not clear why integrated programme planning was dispensed with and why that affected the multi-sectoral cooperation. The first bullet point under Efficiency in the conclusions gives more clues although the higher level responsibility to integrate the sectors seems skipped over as the siloed approach has been shown as a constraint to efficiency in UNICEF evaluations for many years. The first recommendation seems to point to restricting the ECM efforts to small areas, this should be clarified. What will be unique in the pilot areas for testing the cross-sectoral intervention? The second recommendation is successfully launched from the conclusions. For the third recommendation, it is clear that economic stress is a major driver for CM, but it is unclear which agencies already support livelihood development and where they are targeted (e.g., WFP in refugee areas and NGOs in others). How will UNICEF expand its sectors to include livelihood development? Inform the recommendation with understanding of the external context for livelihood development and how the targeting of high ECM areas could be negotiated. (Suggest to number the recommendations in both the text and the table for easier reference.)
ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	The recommendations are useful and the primary intended users appear on Table 2 and effectively note the relevant criteria, the priority and the main recommendation statement. However, the guidance provided is very sparse and general. Overall the key recommendation statements should be bolded and then the suggested implementation steps set out as bulleted or numbered steps. The process to see the recommendation through to its actualization should be clear to the users and the evaluation team should navigate to more firmly ground the recommendations some of which are broadly expressed.
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 process for developing the recommendations is very well described in the introduction to the section (8). Two workshops were held and involved the evaluation reference group and young female activists.
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.	Yes	The primary intended users or duty bearers appear on Table 2 and it effectively notes the relevant criteria, the priority and the main recommendation statement. Recommendations are listed in order of priority.
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	86%	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 opening pages include the date of the report, the names and organizations of the evaluators, the commissioning organization, the table of contents which includes the relevant tables and figures, a list of acronyms and page numbers.
ii	Annexes include: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework (unless included in report body), list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology, information about the evaluator(s), etc.	Partially	The annexes include the TOR, the evaluation matrix, the Theory of Change diagram and the logical framework. The annexes include the qualitative sample in Annex II, however, the names of interviewees are not included. Other annexes offer more detail on analytical processes, such as cost effectiveness and target versus achievement Figures 8 and 9 which are aggregated and not differentiated. Please include the data collection instruments, the list of interviewees and site visits including the names of woredas visited, and move the documents consulted list to the annexes.
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 structure is mainly easy to navigate although readers would benefit from numbered paragraphs. The sections are clearly numbered.(Note: Numbering the paragraphs would improve readability and referencing.)
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 structure accords to UNICEF guidelines. It would benefit readers to have a separate section titled The Evaluation Object outside of the background.
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 report is just slightly over 41 pages without exec sum and annexes and fits the TOR requirement of 40 pages.
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 generally free from grammar, spelling and punctuation errors.
iii	Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled, and referenced in text.	Partially	There is effective use of visual aids to convey information which are clearly labeled. However, a number of Figures are blurry and hard to read, such as Figures 6 and 7 and the numbers reached are not discussed in the text. The Theory of Change diagram in Annex VI is blurry and hard to read. Please sharpen these images.
SECTION I:	**EVALUATION PRINCIPLES (weight 10%)**	88%	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 report has couched CM as a human rights violation and notes that the research methodology follows a human rights based approach. The findings note that Ethiopia has ratified many human rights instruments including CEDAW, the CRC and the African Charter on Human and People's Rights, among others. (Noted in questions 3.2 and 12.1 that this information is more appropriate in the background.)
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	Page 16 notes that two reference groups were established, one of UNICEF Country Office staff and one of young female activists representing the duty bearers and the rights holders. Stakeholders were involved as informants.
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	Language is empowering and inclusive and avoids bias. The terminology of rights holders and duty bearers is used. Data is presented on aggregate numbers of adolescents (Figures 6 and 7) reached but not discussed in the text. The TOR has presented a complete picture of the targeted people both rights holders and duty bearers and their numbers and these numbers are also found on the Logframe Annex VII. However, these are not differentiated in the findings discussions. Suggest to include more detail on the numbers of adolescents referred to in Figures 6 and 7 and whether they are from humanitarian contexts or other.
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.	Yes	Disability and persons living with disability, particularly girls form part of the findings and noted that there is a low number of girls with disabilities included in the programme.
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)	8	

i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Fully integrated	GEEW as part of the ECM approach is fully integrated into the evaluation scope, approach, design, questions and data collection strategies.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	The evaluation has selected a gender responsive methodology, methods and tools in order to gain insight into the causal factors of CM.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	Conclusions and recommendations reflect the strengths and issues in the programme implementation to ECM.

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. 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? b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? d. 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?