GEROS Evaluation Quality Assurance Tool

Version: September 2021

Evaluation of UNICEF Lebanon Education Programme (2016-2022)

REPORT RATING SUMMARY			
Overall Rating		70%	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	Evaluation of UNICEF Lebanon Education Programme (2016-2022)
Report sequence number	Lebanon/59/2023/19190
Region	MENA
Year of report	2023
Office	Lebanon
Coverage (countries)	Lebanon
ToRs present	Yes
Date of review (dd/mmm/yyyy)	January 10, 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	Summative and formative
Evaluation strategy	Mixed methods
Evaluation design *(primary method used)*	Theory-based
Evaluation level	Output & Outcome
Geographic scope	National
Primary SDG(s) covered *(number)*	4, 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 satisfactory report that may be used by decision makers pending revisions. The complexity of the scope of the evaluation and the wide range of evaluation questions is well noted and the strong set of findings very useful for future programming.

The report offers the following strengths:
- The executive summary is useful for informing decision making and does not introduce new information.
- The context of the intervention is clearly described and includes the major issues affecting education and children living in Lebanon related to the implementation of the intervention.
- The intervention is clearly described, with outputs and outcomes, and the main Lebanon Country Office – Education Programme (LCO-EP) counterparts are mentioned, reflected in the evaluation matrix, and connection to SDG4 brought in.
- The purposes and objectives of the evaluation are clearly noted as well as temporal and geographic scope. The evaluation questions are appropriate for meeting the objectives and purpose.
- The methodology is strong to test a theory-based approach and the limitations clearly mentioned with mitigation measures.
- The ethical protocols are very effectively covered in Annex 6.
- The Ripple Effects Mapping (REM) has potential to be an innovative practice as used in this example. A good practice is highlighted in the Table 5 which clearly notes the achievements against the plans.
- The findings systematically address the questions and use credible evidence and triangulation, and in most cases strongly present the causal factors.
- The conclusions are effectively separated by themes and the recommendations by systems and programme. There are four bolded lessons which are generically applicable.
- The recommendations are useful and actionable.
- The report structure includes the relevant information and is clearly written.
- There is strong rights basis and the findings draw in the relationship to the humanitarian situation as well as effects of the programme on the perceptions of the refugees and Lebanese.
- The UNSWAP criteria are satisfactorily integrated in the approach, findings and recommendations.

Areas for improvement include the following:
- The report is missing some information which is needed to fulfil the QA criteria (including sampling, methodology, limitations, ethical standards, among others) and most of this is found in the inception report but not brought into the final report. Most readers will not refer to the inception report, thus the gaps need to be addressed.
- The executive summary needs to be revised based on the changes recommended for the conclusions, and recommendations. It does not include the purpose, objectives and

•The executive summary needs to be revised based on the changes recommended for the conclusions, and recommendations. It does not include the purpose, objectives and users. It is 8 pages whereas 5 are recommended.
•The object of the evaluation (section 1.3) does not cover the budget or provide an approximation, and the details are found later in the findings. There is no complete mapping of the stakeholders with their roles in the evaluation and the linkages between them.
•The scope of the evaluation (section 1.3) has to include the extension of the dates of the evaluation to 2022, as different from that stated in the TOR, and the limitation of the scope to Formal Education (FE) and Non-Formal Education (NFE) covered by another assessment clearly explained also as different from the TOR.
•The process to finalize the evaluation questions is not fully explained and noted where variations were made from the TOR.
•As this is a theory-based evaluation, the Theory of Change (ToC) as represented in the results framework and the reconstructed ToC should be a continuous thread throughout the report, whereas it is not mentioned past section 2.0 or brought into the findings.
•Some conclusions include findings and recommendations and some lack the needed clout to launch the recommendations.
•The lessons are strongly critical of the national system and need to be revised to remove language which may be perceived as offensive.
•The recommendations for system and programme also contain findings and have different formatting and need to be more systematically organized.
•Some required annexes are missing such as the Theory of Change diagram, the list of interviewees.
•The body of the report is 80 pages, 20 over the TOR recommended length.
•An evaluation reference group is not mentioned and its role.

Recommendations for Improvement: *The rater will identify topline recommendations to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, resources will be cited to assist evaluation managers in overseeing future evaluations.*

The following are priority recommendations with others to be found in the review.

1.Revisit the conclusions (Section 4.2) and repeat the purpose and objectives in the introduction to the section. Add clout to the conclusive statements by tightening the language to reflect the higher level of analysis, and remove findings, with the intention of focusing on the conclusions that will launch the recommendations, possibly taking some out or combining. Number them and refer to the conclusions next to the relevant recommendation. Some suggestions are included in the review for tightening wording.
2.Rewrite the lessons (Section 4.1) to remove any language which may be perceived as offensive to the government partners as they will also be recipients of the lessons.
3.Combine the system and programmatic recommendations (Section 4.3) on a matrix or create separate matrices with the same formatting, for easier reference, noting the appropriate conclusions. Add more clout to the recommendations by tying them to conclusions, removing findings and other substantiations and clarifying the steps for implementation, and the users. In the introduction paragraph to the recommendations, describe the process to gain buy-in for the recommendations including the role of the evaluation reference group.
4.Revise the executive summary based on the changes suggested for the conclusions and recommendations and ensure a high level of readability for decision makers. Include the evaluation purpose, objectives and intended audience.
5.Information which is needed to fulfil the QA criteria (including sampling, methodology, limitations, ethical standards, among others) should be brought in from the inception report possibly as annexes and summarized in the main text with reference to the annex.
6.The object programme (section 1.3) should include an approximation of the budget and a complete mapping of the stakeholders with their roles in the evaluation and the linkages between them.
7.The scope of the evaluation (section 1.3) has to include the extension of the dates of the evaluation to 2022, and the limitation of the scope to Formal Education (FE) and Non-Formal Education (NFE) covered by another assessment clearly explained.
8.Include a description in the methodology section (Section 2.0) of the process to finalize the evaluation questions and note where variations were made from the TOR as well as incorporating the Theory of Change which was reconstructed by the team and referring to it where possible in the findings, such as where important gaps are found and need to be addressed in a revised ToC.
9.Reduce the pages in the main report and the executive summary to closer to the recommended and gain agreement from evaluation management if the page limits need to be extended. Some suggestions are made in the review for shortening.

SECTION RATINGS

SECTION A:	**EXECUTIVE SUMMARY (weight 5%)**	50%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
i	Is clearly presented, serves as a standalone document useful for informing decision making, and is of relevant conciseness and depth for key users (Maximum of 5 pages unless otherwise specified in ToR).	Partially	The executive summary is useful for informing decision making. However, it is too long at 8 pages and the length is a constraint for readers who need a concise summary. The recommendations are not unified and written in different formats for system and programme. Suggest to limit the pages to a maximum of five by consolidating the recommendations such as by combining them on a matrix and eliminating all of the excess verbiage including findings that executive readers do not need. The programme level recommendations matrix presented is useful and effectively clarifies the criteria, the priority and time frame. This structure could be used to combine the two, however, the findings are mixed in and contribute to the excess text. Suggest to remove the findings statements from both sections.
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 an overview of the intervention (please spell out RACE), the methodology, key conclusions and recommendations. The evaluation purpose, objectives and intended audience are not included. Suggest to add a short section or paragraph prior to the overview of the intervention or the methodology.
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, some significant information needed to understand the evaluation is missing, as noted in Question 1 iii.
SECTION B:	**BACKGROUND (weight 5%)**	71%	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 1.2 presents the description of the intervention, Lebanon Education Programme (LCO-EP), including the rationale, the three pillars and their outcomes of the UNICEF education programme, as well as the evolution of the programme over the years it is being evaluated. The total budget devoted is unclear; while the source of funding appears to be the Transition Resilience Education Fund (TREF) which kicks in by 2021 and the budget designated by the Ministry of Education and Higher Education (MEHE) in 2010, the funding during the programme period being evaluated is not discussed. A detailed breakdown of funds is not found until the Findings - Efficiency section where they are clearly graphed. Suggest to clarify the aggregated funds in the background section.
ii	Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (i.e., institutions/organizations; communities; individuals…), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cites, sub-regions…) and in terms of numbers reached, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).	Partially	The discussion of stakeholders as groups and individuals is scattered among the discussion of the UNICEF education programme in section 1.2 but no focused discussion or mapping of stakeholders is found in the body or the annexes which is typically developed in the inception phase. It is unusual that the inception report does not include a stakeholder map or matrix which classifies and characterizes the various rightsholders and duty bearers and their locations, while the TOR offers a chart in part 29 that could be expanded to include the rights holders. It will be important to include some form of stakeholder mapping that consolidates the stakeholder picture for the readers of the evaluation report. The numbers reached are also scattered and numbers of teachers and other duty bearers are unclear. Also unclear are the numbers of the students and the most vulnerable in Lebanon. Please develop a matrix that shows these groups, their numbers or organizations and location and numbers by gender, age and other factors such as disability. Refer to other evaluations with stakeholder matrices or stakeholder analyses.
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 Introduction Section 1.0 is devoted to the background including the Syrian war, the economic downturn, the Beirut blast and COVID-19 and their impact on the education system in Lebanon. The characteristics of the Lebanese education system are also discussed and the situation of the Syrian refugees in the system. The Reaching All Children with Education (RACE) strategy was developed by the MEHE in 2014 to address the Syrian refugee crisis, and due to the fact that the refugee children outnumbered the Lebanese. RACE II was implemented during the time under evaluation. The context relationship to the intervention is described in the rationale, section 1.2.1.
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Yes	Section 1.2.1 talks about SDG 4 in relation to the programme and SDG 4 is brought in through the evaluation matrix.
iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	The status of beneficiaries is discussed in Sections under the Background outlining 1) the effects of the war, the economic downturn, the Beirut blast and COVID-19 on children and their families, including refugees mainly from Syria, 2) on the disruption of education activities, along with 3) the impact of the decrease in purchasing power on teachers and gender parity among students. The section 1.1.5 on characteristics of the Lebanese education system identifies the main issue, including quality, physical capacity, and language issues for refugees, and describes efforts via RACE I to enrol out-of-school Syrian children in formal education and language training. The RACE II strategy (2017-2021) aimed for greater access, inclusion, and structural and teaching capacity. The section on 1.2 description of the intervention outlines further issues, such as increases in number of drop outs and of child labor as well as a high rate of non-attendance (30%) due to high costs of materials and transport, among others.
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	The main counterparts are mentioned in section 1.2.2 including international agencies, national agencies, the general category of NGOs, and parents and children. Please identify linkages between the implementing agencies, the development partners, the rights holder, and duty bearers such as teachers and any other relevant stakeholders. A mapping of stakeholders is useful in this regard or combining linkages into the stakeholder matrix. The roles and programmes of the other development partners in education such as UNESCO, UNHCR, the World Bank and relevant NGOs and CSOs should be included. It is also helpful to include numbers such as numbers of students in school and out of school, teachers, technical and administrators to offer a more complete picture of the education system.

ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	A generalization of the roles of the stakeholders, such as supporting UNICEF and the government, is made in section 1.2.2 but identification of the specific contributions or the roles are not consolidated. Please include a stakeholder matrix that describes the contributions of each of the stakeholders and their roles in the programme including specifying UNICEF's roles such as for advocacy.
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.	Yes	The purpose, scope and objectives are defined in section 1.3. The purposes are to assess results (outputs, outcomes and impacts) and the learning objective is stressed. (Use the past tense rather than "will be") A chart depicts the dimensions for assessment and learning that are important in the data collection. Section 1.3.6 describes the evaluation users and that the report will be available in Arabic.
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	The objectives and scope are clear. There are 5 specific objectives which adhere to those of the TOR, including reconstruction of the Theory of Change and identification of good practices and gaps. The first limitation in section 2.2 noting that different types of non-formal education (define NFE) were supported by the programme is very important in the context of Lebanon and the refugee situation. The fact that NFE is not covered in detail needs to be noted in the purpose and objectives.
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).	Partially	Section 1.3.5 covers the temporal and geographic scope. The UNICEF support for the education programme in the entire country is included although sampling of schools was modified to account for safety and security concerns. The evaluation covers 2016-2022 and it should be noted that 2022 was added in the inception phase and differs from the TOR (2016-2021). The first limitation that different types of non-formal education (define NFE) were supported by the programme is very important in the context of Lebanon and the refugee situation. The fact that NFE is not covered in detail needs to be further noted in the scope section, outlining what will and will not be covered as agreed by management and the team.
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.	Yes	Section 1.2.2 describes the intended results and presents the three associated pillars. Each pillar includes outcomes which are described as well as a general description of the inputs and outputs planned to achieve them. The evaluation is testing these as per questions on the evaluation matrix. (Please clarify that the intended results are all being tested.)
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	Section 1.2.2 describes the 3 pillars (outcomes) of the intervention; the inputs and outputs are described in a narrative form. An evaluation matrix is presented in Annex 1 and contains the questions, indicators, data collection methods, data sources and data analysis methods.
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 Terms of Reference requested that the Theory of Change be reconstructed by the evaluation team. The Inception report presents a reconstructed ToC in Figure 1 but this is not brought into the final report. Since the assessment of the ToC was a main objective of this theory based evaluation, it must be brought in either in the text or the annexes and referred to in the text, as most readers will not have access to or not read the inception report. An explanation of the assessment of the ToC by the team should be there at least in a few sentences, such as whether there was any adjustment by the team or the theory and causal relationships as discussed in the text is simply represented in graphic form.
SECTION D:	**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	64%	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 Section 2 includes Figure 2 which presents short descriptions of the tasks that were part of the evaluation approach in the inception, data collection and data analysis phases. According to the evaluation matrix, the criteria and questions are appropriate, however there is no explanation regarding the way the questions were finalized such as taken from the TOR or developed further by the evaluation team. The inception report under section 2 (Overview of the progress to date) offers more information on Task 3 - evaluation matrix validation workshop where questions and matrix were vetted with stakeholders. This is a critical feature of the process of designing the evaluation and should be brought into discussion in the methodology section as readers will likely not read the inception report. A description of this process should appear in section 2.1 to inform the readers how the questions were selected and developed and importantly that they were vetted with stakeholders, which contributes to the validity of the findings, conclusions, and recommendations. The Effectiveness question in the TOR which pertains directly to the ToC: "To what extent has the Theory of Change (ToC) been followed in implementation of programmes to achieve results identified in the Programme Strategy Notes (PSN)?" has been reworded or not included. It is noted that the number of questions is high and contributed to the length of the report as per Question21 i.
ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.	Yes	The evaluation matrix is well developed with questions, indicators, data collection methods, data sources and data analysis.
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	Section 2.1 Overview of the methodology focuses mainly on data collection and some information is provided on sampling criteria. The inception report more fully describes the approach and set of methods and suggest that the information is included as an annex and referred to in the main text and/or a short summary is included in the text.
ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Yes	The data sources are appropriate and include both qualitative and quantitative sources, including desk research, interviews, observations and deep dives.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Partially	Sampling is mentioned in Section 2.1 with a list of those actually interviewed, however, the sampling strategy is not fully presented. Table 2 in the inception report summarizes the proposed sampling of central level stakeholders, although it is unclear if it includes sub-national. The inception report does a good job of indicating that diverse perspectives were sought and notes careful consideration of how the sample will be used to complement other data. Suggest that the Table 2 is updated to show the interviews that actually took place against the planned, specifying the organizations and include the sub-national level and that this table is placed in the text or the annexes with an explanation appearing in the main text.
iv	Clear and complete description of the methods of analysis.	Partially	There is sparse description of the methods of analysis in the methodology section (2), such as Ripple Effects Mapping (REM) which might be considered an innovative practice. Suggest that the strong discussion of the data collection and analysis methods found in the Inception report section 3.4 is added as an annex ahead of the tools in Annex 7, summarized and referred to in the main report.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Partially	As per section 2.1 this is a theory driven evaluation based on the Theory of Change, therefore it should be mentioned that the team strategy is to track and draw the causal connections between outputs and outcomes indicating whether results have been reached or will be reached. The "moments of change" approach taken by the team is important, however, this analysis should be tied to the Theory of Change and verifying whether the ToC is valid or whether it should be altered as per the findings. Suggest some rewording and strengthening the ToC approach. Section 3.1 in the inception report presents a somewhat stronger explanation.

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	The limitations are described in section 2.2. The first limitation that while different types of non-formal education (define NFE) were supported by the programme, NFE is not covered through data collection in detail. Rather the NFE is covered by another assessment and those results considered. This needs to be noted in the scope section, (what will be covered and not covered) given the importance of NFE in the context of Lebanon and the refugee situation. Further, the reasons for non-coverage, which should be vetted in the evaluation design phase, need to be better outlined as the intention to reach the most vulnerable is critical for UNICEF. The TOR states: *"UNICEF also supports non-formal education programmes, as part of the multiple flexible pathways, with the support of implementing partners. These programmes target out-of-school children aged 3 to 14 in line with MEHE's 5 year-plan."* Suggest to strengthen the reasoning and refer to the discussion in the inception report that the scope was too large for the resources devoted to the evaluation. As such, it was agreed to limit the scope and the NFE would be covered by a different assessment and those results considered. Suggest to place each limitation in a matrix with a column indicating the mitigation measure taken by the team. Currently more than one limitation appears under one bolded statement (i.e. limitation 2 and 4) on accessing the most vulnerable.
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).	Partially	Annex 6 effectively covers the ethical protocols followed by the team and UNEG principles are briefly mentioned. The TOR notes that Ethical Considerations should form part of the order of the report but they are missing except for the inclusion paragraph. Suggest that the discussion in the inception report on Ethical Standards (3.2) is brought in which features a very good Figure 5 demonstrating the UNEG standards. The main text may feature a summary and the entire section 3.2 become part of the Annexes.
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 6 effectively covers the ethical protocols followed by the team including stakeholder participation, informed consent, safety and protecting privacy.
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.	Partially	There is evidence that the evaluation used innovation in use of the Ripple Effects Mapping (REM) which can collect untold stories and capture wider impact and is a technique which is publicized on the internet. To help others include this technique, its use in this evaluation could be further described as to what findings it contributed to and possibly appear as a box or in the annexes. The description in the inception report on Task 9 is useful but it is still a theoretical concept. Suggest the team further describes what insights this technique supported and details on its use.
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.	Yes	The findings section 3.0 follows the questions systematically and indicates a strong level of evidence used to answer each question. The main discussion is balanced with evidence from duty bearers presented in boxes.
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Partially	While the findings discussions bring in some aspects of the Theory of Change, they are generally not referred to as such. After Section 2.0 the ToC is not mentioned. The Effectiveness question in the TOR which pertains directly to the ToC: "To what extent has the Theory of Change (ToC) been followed in implementation of programmes to achieve results identified in the Programme Strategy Notes (PSN)?" has been reworded and the findings do not directly refer to the ToC but do note whether the assumptions were followed. Suggest the team review the findings and mention the relevant part of the Theory of Change as it pertains to the finding in accordance with the theory based approach.
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 bring in background, secondary and primary data, documentation, as well as anecdotal information. There is evidence of triangulation in the analysis.

ii	Findings are clearly supported by, and respond to, the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	The findings are clearly supported by the evidence presented, and include both positive and negative findings. Table 5 presents the status of implementation against the Programme Strategy Notes and mentions the sources of data. This is an example of good practice to clarify the analysis. The findings of the separate assessment of the NFE are also included.
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).	Partially	While the findings are very well developed, in some cases they stop short of getting to why the logical chain planned in the Theory of Change is not always clear and what the causes are. For example, as it turns out the NFE issues are important as to why there remains a high level of out of school children. It also reveals that the national tensions among the Lebanese and Syrians is a factor. There are some successes noted such as better integration and would be interesting to note what is working in favor of the integration. There may be an intermediate step needed between the inputs, outputs and outcomes which involves overcoming assumptions and bias in the system such as by teachers in order for the education and inclusion goals can proceed.
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	Question 3.3.5 addresses the monitoring system and a conclusion and recommendation is devoted to improvement of the monitoring framework. Various aspects of monitoring appear in the Effectiveness and Efficiency sections of the findings. Noted on Table 5 that the incentive-based teacher performance monitoring system was partially implemented and other implementation aspects of the monitoring system specifically noted. Noted also is that the data collection instruments reflect the monitoring questions.
SECTION F:	**EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)**	75%	Comments on Rating
Question 15.	Do the conclusions clearly present an objective overall assessment of the intervention?		
i	Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).	Partially	The conclusions (Section 4.2) are effectively separated by themes, such as overarching, access to education, quality of education, and system strengthening. They are forward looking for future programmes. Suggest that the purposes and objectives of the evaluation are repeated in an introductory paragraph to satisfy readers that they have been respected. In this vein, the importance of the Theory of Change should be emphasized. Care should be taken to avoid repeating findings and already inserting recommendations, but rather setting the stage for them by stating the problem. For an example of tightening and remaining conclusive: "In light of stakeholder feedback, an aspect that has not been tackled either, but should, is the humanitarian nature of many access interventions in the programme. These interventions rely on massive financial transfers. The view among the interviewed stakeholders is that there should be more strategic thinking about emergency interventions. While the latter are still needed, the balance should tip more towards development work to achieve better sustainability in the long run." could instead read: "Humanitarian access is weakly strategically addressed in view of the massive financial transfers involved and does not contribute widely to long term sustainable development, as the humanitarian to development continuum recommends." Further, given there are many conclusions, suggest they are numbered and then associated with the relevant recommendation(s).
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	Conclusions are derived appropriately from findings and present both strengths and limitations. Although there are many examples of insight, further tightening of the language will give more clout and reflect the higher level of analysis. This is a critical conclusion (4.2.3): "The fieldwork suggests that teachers and principals could benefit from more emphasis within the program. The productivity allowance under TREF responds to a significant need for financial support. However, there is also a need to develop non-financial interventions to build teachers' child rights-based attitudes, as well as maintain their motivation and engagement. These could also serve to ensure teacher well-being in the face of multiple pressures." Suggest to add more clout: "The addition of curriculum reforms and a TREF supported productivity allowance has improved teaching quality, however there is not enough attention to supporting teacher and principals capacity in upholding child rights and to maintaining their motivation and engagement."
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	Four bolded lessons learned are offered (4.1) and they have wider applicability and relevance.

ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	The lessons in bold are clearly presented but are strongly critical to the national system and this is not usually the intention of lessons. Suggest to avoid words such as "dubious" in regard to government capacity and use more diplomatic language. It is also not necessary to cite donor input as this may also be harmful rather noting that the lack of reliable data was an issue. Suggest that the first three lessons and their details are revisited and downgrade criticisms, such as replacing "lacking" with weak capacity or capacity that requires strengthening, along with greater recognition that the lessons are not one sided but are applicable to all management stakeholders.
SECTION G:	**RECOMMENDATIONS (weight 15%)**	50%	Comments on Rating
Question 17.	Are recommendations well grounded in the evaluation?		
i	Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.	Partially	There are two groups of recommendations (system and programme) in section 4.3 and they are based on the findings and presumably the conclusions. However, they are presented in two different formats suggesting they are not uniformly developed. (Both will require a management response.) Both sections include findings which burden the text and are not needed; these should be removed and the recommendation strongly emanating from the conclusion and suggest that the numbered conclusion(s) be attached to the recommendation. Suggest that all recommendations be placed on a matrix or matrices and follow the same format. Taking the system level recommendations for example, the first recommendation is emanating from the first and second paragraphs under 4.2.4 and concluding (but not using very strong language which can be improved and removal of the finding and source) that there is not enough attention to the policies and reform agenda
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 but some contain justifications which should be apparent in the findings and conclusions. Recommendations should be numbered and matched with the appropriate conclusion. For the first recommendation under system - the "well placement" of UNICEF is extraneous but rather that UNICEF should make good use of its experience and availability of resources. The steps suggested are appropriate and actionable. It is unclear under the system section which are the key recommendations. Suggest to create a matrix as found under the programme or join them together and clearly stipulate the key recommendation and then the steps to make it actionable. For the programme recommendation, for example, the first one, the finding is not needed but rather to describe how supplies will be made more relevant, and what needs to be done.
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).	Partially	The process of developing the recommendations is not described in this section although the task list in Figure 2 implies that there was stakeholder feedback. Please include in the introduction to the recommendation section a description of the process for developing the recommendations and how the duty bearers were involved or not involved and why. The role of the evaluation reference group should be noted as well.
Question 18.	Are recommendations clearly presented?		
i	Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.	Partially	The systems recommendations are directed at UNICEF but a more precise identification would be very useful, such as the education section, or coordination bodies. The programme recommendations are clearly prioritized but contain too many findings, suggest removing these or reformulate the statements to be what it is suggested to do: such as where digital equipment is underutilized, what to do to make it more useful.
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	79%	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 organization of the evaluators, the commissioning organization, the Table of Contents with tables, figures and annexes, the background, description of the intervention, as well the list of acronyms.
ii	Annexes include: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework (unless included in report body), list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology, information about the evaluator(s), etc.	Partially	The annexes include the TOR, evaluation matrix, data collection instruments, documentary evidence. The list of interviewees, the Theory of Change and its explanation should be included. It is suggested that the annexes include the expanded methodology as found in the inception report and other details to round out the picture of the approach and required elements as mentioned such as the ethical standards and the sampling approach. Many of these are included in the inception report and can be transferred and updated.
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 easy to navigate with numbered sections and clear titles and subtitles. (Note: As per the comments in Q17 I and Q17 ii, navigation of the recommendations would improve with numbering them.)
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.
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.	Partially	The body of the report is nearly 80 pages excluding the executive summary and annexes. This is considerably longer than the 60 pages specified in the TOR and calls for a considerable reduction in page length. However, it is to be recognized that there are many evaluation questions and the answers to all of the questions are included. The evaluation management should decide whether this amount of detail is required and correspondingly, a longer page length. Suggest that the details found in the background are reduced to a summary and the longer versions placed in the annexes.
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 there are very few 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	The infographics convey key information and are labeled and referenced, but are notably missing the Theory of Change diagram which needs to be included.
SECTION I:	**EVALUATION PRINCIPLES (weight 10%)**	76%	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 did a good job of reference to the rights basis and this is mainly found in the findings on relevance. Further the report explores the connection with the humanitarian situation.
ii	Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).	Partially	The level of participation of key rights holders and duty bearers is addressed in the tasks described. There is no mention of a reference group in the final report as stipulated in the TOR: (32) Evaluation Reference Group (ERG) "A reference group will be formed to review, guide, and endorse the deliverables and ensure that the evaluation answers all questions. The ERG will consist of members including UNICEF Education staff, key donors, and MEHE. The reference group will be entrusted to guide the evaluation process, including by providing strategic inputs across the whole process, from the design phase to the delivery and comment on the final report. It will also be involved in the recommendation co-creation workshops and in the dissemination." The formation of this group and the members participation is described in the inception report under Tasks 12 and 14 but not brought into the final report. The ERG is a critical component of the evaluation process that helps to reduce bias and offers important advice. Please bring this into the appropriate sections or devote a short section to the description of who is in the ERG and the roles of the ERG or in the annexes.
iii	Language is empowering and inclusive, avoiding gender, heterosexual, age, cultural and religious bias, among others; use terminology of rights holders and duty bearers; data is disaggregated by marginalized group; differential results are assessed (distribution of results across different groups).	Partially	The language is mainly empowering and inclusive except where noted in the lessons, where it needs to be modified to be more tactful. The terminology of rights holders and duty bearers is not used and suggest to incorporate this language where appropriate. The findings note the different groups and the results across groups.
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	Section 2.1 includes discussion of the use of the human rights based approach and attention to race, gender, nationality, disability and other factors. The findings point to lack of disability statistics, and UNICEF support to the MEHE on disability inclusive policies. Since the premise of the programme is to leave no child behind, the findings address this.
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)	7	
i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Satisfactorily integrated	The report mentions gender in terms of inequalities in the background noting that the high gender equality gap contributes to the gaps in education; the Theory of Change outputs target reduction of gender barriers; and the equality and inclusion lens used by the evaluation. The integration of GEEW data is assured through two specific evaluation questions. However, there is no evaluation objective that specifically addressed gender or notes whether girls are among the most vulnerable.

ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	An effectiveness question was devoted to gender transformation; gender is addressed in the relevance findings and issues in gender disaggregation of data. The data collection tools reflect this focus. Section 3.2.6 addresses the extent to which the programme has been gender responsive or transformative. Section 3.2.6 also answers the question as to how well the data was gender disaggregated.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	The findings, conclusions and recommendations reflect a gender analysis and note connections to gender-based violence. The relevance finding notes the national policies and capacity issues to produce gender related data; the Teacher Training Curriculum model (TTCM) included a module on gender mainstreaming. While there are overall issues in the country with gender disparity, it seems that at the education level there is purported to be gender parity in schools but the picture of how girls are affected who are out of school or frequently absent is fragmented in the sections. The report covers unexpected effects in section 3.2.3 and this includes UNICEF support for government decisions to place Syrian children in second shifts with limitations on activities compared to the first shifts and which has caused some controversy. In addition, the perception of UNICEF support mainly directed to Syrian children has caused controversy among Lebanese.

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?