

GEROS Evaluation Quality Assurance Tool

Version: September 2021

Summative Evaluation of Global Partnership for Education Programme Puntland

REPORT RATING SUMMARY		
Overall Rating	54%	Fair
●●●●●	Exceptional (96% - 100%)	5
●●●●●	Highly Satisfactory (87.5% - 95.99%)	4
●●●●●	Satisfactory (62.5% - 87.49%)	3
●●●●●	Fair (35% - 62.49%)	2
●●●●●	Unsatisfactory (0% - 34.99%)	1
Meets UNICEF/UNEG standards for evaluation reports in some regards, but not all. Decision makers may continue to use the evaluation with caution, but substantive improvements are possible.		
REPORT DETAILS		
Title of the evaluation report	Summative Evaluation of Global Partnership for Education Programme Puntland	
Report sequence number	Somalia/63/2024/2012	
Region	ESAR	
Year of report	Dec-24	
Office	UNICEF Somalia	
Coverage (countries)	SOMALIA	
ToRs present	Yes	
Date of review (dd/mmm/yyyy)	February 7, 2025	
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	No	
Every child learns	Yes	
Every child is protected from violence and exploitation	No	
Every child lives in a safe and clean environment	No	
Every child has an equitable chance in life	No	
Gender equality (cross-cutting)	Yes	
Humanitarian action (cross-cutting)	No	
Evaluation object	Programme	
Evaluation type	Summative	
Evaluation strategy	Mixed methods	
Evaluation design (primary method used)	Theory-based	
Evaluation level	Output & Outcome	
Geographic scope	National	
Primary SDG(s) covered (number)	4	
EQA Summary: <i>The rater will provide top line issues for this evaluation relevant for feedback to senior management (positive and negative), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight best practice/added value elements and the level of complexity of the evaluation.</i>		
<ul style="list-style-type: none"> ◆The evaluation report is, overall, well written, and well presented. It is a fairly easy read, despite its length, at just over 60 pages. ◆The context section is mainly copied from the UNICEF ToRs, no new information is presented, and readers may have benefited from understanding more about how the context affected the education sector, particularly in Puntland, and particularly for girls, people with disabilities, and minority groups. ◆Similarly, the description of the intervention is summarised from the ToR. Again, readers may have benefited from a more full and coherent description of the intervention ◆The evaluation was ambitious; the purpose was to cover both accountability and learning, with a strong emphasis on gender and inclusive education. The ambition is also reflected in the number of evaluation questions, at over 30. ◆Methodologically, the evaluation presented a strong grounding, applying appropriate methods. However, there are some weaknesses with the methodological approach, such as not sufficiently using the theory of change. ◆Guided by the purpose and the evaluation questions, the methodology for the evaluation presented a strong alignment with gender and social inclusion. This was for both the approach of the evaluation, as well as what and how the evaluation assessed the intervention. This is reflected in the findings section which offers a diverse and inclusive range of perspectives. ◆The conclusions, lessons learned and recommendations sections are quite strong, but with small omissions which could easily be resolved – for example, referring to targets in the conclusions, and targeting the recommendations at specific stakeholders. 		
Recommendations for Improvement: <i>The rater will identify topline recommendations to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, resources will be cited to assist evaluation managers in overseeing future evaluations.</i>		

- ◆The evaluation team may have considered providing a short introduction before the context section. This 'set up' the context section, providing the reader with an understanding of why context is discussed.
- ◆The context section may have offered specific information on the how context is affecting thematic area (i.e. how the context in Puntland is specifically affecting education and education for girls, children with disabilities and minority groups). This could have been developed from the evaluation team's research and understanding, rather than just relying on the ToR.
- ◆The evaluation team may have provided more description on the intervention. For example, a clearer description of the objectives of the intervention and the 'theory', the hypothesis of how the intervention activities will lead to change. Developing and presenting a theory of change would not only help the reader understand what is being evaluated, but it would also have provided a strong framework for the evaluation approach itself.
- ◆As well as describing the intervention, it is recommended that the evaluation team present the stakeholder analysis earlier in the report, so that the reader understands who the rightsholders/beneficiaries are and who are the duty bearers.
- ◆It is also recommended that the evaluation team try to negotiate fewer evaluation questions. This can often be challenging, as commissioners have lots of things they want to learn about their interventions. However, the large number of questions is challenging to answer in a concise report, with limited resources for the evaluation process.
- ◆Along with the stronger focus on a theory-based evaluation approach (using the theory of change), the evaluation team may have considered providing more description about the use of 'non-target' schools to compare how the intervention/target schools changed.
- ◆For the conclusion, it is recommended that the evaluation team further consider the purpose of the evaluation; reflecting on the achievements compared with the targets of the intervention.
- ◆Finally, it is recommended that the recommendations are targeted, articulating who is best placed to implement and take forward the recommendation.

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 mostly well written and serves quite well a standalone document. It is of good length, being 5 pages and contains all critical information from the main body of the report. There are some minor formatting issues, such as large spaces in the recommendations section, and the order of the executive summary is different to that of the main report – with the recommendations coming before the conclusions. The executive summary may have benefited from an edit, removing some of the unnecessary statistical information - instead this could have been a summarised narrative, and particularly editing out unnecessary statistical information on effect sizes and standard deviations.
	ii Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key conclusions on findings, lessons learned if requested, key recommendations) as per the ToR.	Partially	The executive summary covers most of the required information, only omitting a summary of the lessons learned section. As in the main report, the Overview of the Intervention section would benefit from providing information on the duty bearers and rightsholders – but these are provided later in the evaluation objective section.
	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	As reported above, and as in the main body of the report, the executive summary would have benefited from providing a description of the rightsholders and duty bearers of the intervention. The reader has to read on to the next section to understand these stakeholders. No new information is presented however.
SECTION B:	BACKGROUND (weight 5%)	29%	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 intervention is taken directly from the UNICEF ToR. The dates of the programme are provided and offer an indication of the implementation status, and there is a clear table (again taken from the ToR) that presents the costs of the programme, made up from three GPEP grants. This short description goes some way towards being a suitable introduction to the intervention, though it would have benefited from more description, such as management arrangements, further description of intended results and intervention approach. Additionally, although the description is taken from the UNICEF ToR, the order is slightly different, and therefore some of the acronyms have not been introduced before they are used. For example, the three grants that make up the intervention (ESPIG, MCA, and AF) are used in the first paragraph of Section 2.0 but have not been introduced beforehand.
	ii Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cities, sub-regions...) and in terms of numbers reached, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).	No	Again, there is a short description of the rightsholders that has been taken from the ToR. This offers some narrative breakdown of rightsholders. There is no description of duty bearers. This section may have been greatly strengthened by more description of the rightsholders. For example, number of children targeted by the intervention, ideally broken down by rural, urban, coastal; by gender; an estimated number of targeted children with disabilities; as well as number of targeted schools (similarly disaggregated by urban, rural, coastal etc). Additionally, some information on the duty bearers – the government ministries responsible, implementing partners, the role of UNICEF in implementation, any non-state actors and the role of any community groups – would have helped the reader understand the stakeholders involved in the intervention.
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	The report presents an introduction to the national and regional context of the intervention. The description covers socio-economic, policy and political issues, all of which are linked to the education sector, the main subject of the intervention. However, much of the narrative is directly copied or slightly summarised from the UNICEF ToRs. This section would have greatly benefited from more in-depth knowledge from the consultant team rather than from the UNICEF ToR. Developing a good understanding of the context is an essential first step for the evaluation team when carrying out an evaluation. Presenting this understanding at the start of the report, in the context section, helps the reader 'place' the intervention and the evaluation.
	ii Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	No	The contextual background section does not offer any linkages or reference to the SDG. This may have been one area where the team could have easily expanded upon the ToRs. Illustrating how the intervention supported the advancement of SDGs, particularly highlighting how the intervention advanced SDG 4, would have helped set the intervention in a good context.
	iii Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Partially	The context section is presented before the reader has any understanding of the intervention. Therefore, when reading this section, it is not directly clear who are the rightsholders/beneficiaries. However, the reader can deduce that the education in Somalia is the subject of the intervention and thus school age children are the rightsholders. And with this, general, basic information on children's levels of access to education in Somalia is presented – all taken from the UNICEF ToR. The context section would have benefited from introducing the rights holders more clearly. The evaluation team may have considered presenting some statistics on the rightsholder population. For example, the total number of school aged children in Puntland, the number of schools in the state, and a breakdown of some education statistics of the state, such as drop out and graduation rates. However, having said this, it is recognised that data availability may have been an issue.
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).	No	When describing the evaluation purpose, the evaluation report states that the evaluation provided accountability to key stakeholders. The report then lists in the same sentence, the different stakeholders by type. For example, it states the primary users (MoEHE), the donor, the Grant Agent, the implementing agencies. This section would have greatly benefited from a clear identification and description of the stakeholders (implementing agency(ies), development partners, rightsholders, and duty bearers) involved in the intervention. This may have been presented in a table, naming each of the main stakeholders, providing a description of their role, and the relationship they have with each other.
	ii Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	As above, there is no clear identification of intervention stakeholders. Therefore, there is no description of the contribution of stakeholders and their role in the project. As suggested above, the evaluation team may have considered providing this information in a simple table, with rows that present each stakeholder, and columns that describe a) the roles and responsibilities, and b) relationships between stakeholders. Further below in the report, under the evaluation design section, the report provides a stakeholder analysis. This analysis is presented in a table (Table 4), and this table summaries key users of the evaluation and their intended use of findings. This is a helpful table and is the best place in the report to identify some of the stakeholders in the evaluation. The information in this table could have been used and expanded upon (to include rightsholders) to identify stakeholders.
SECTION C:	EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	33%	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	The evaluation purpose is outlined, following the ToR, and states that the purpose is both accountability and learning. The description of the purpose would have benefited from clarifying whether accountability or learning was a priority and then describing a bit more about that purpose. The current description in the report states that the users of the evaluation are just Puntland Ministry of Education and Higher Education (MoEHE). Further below in the report, under the evaluation design section, the report provides a stakeholder analysis. This analysis is presented in a table (Table 4), and this table summaries key users of the evaluation and their intended use of findings. This is a helpful table and is the best place in the report to identify stakeholders and users of the evaluation. It is recommended that this table is moved to Section 3 or Section 4, where the purpose and objective of the evaluation is presented.
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	As stated in the report, the evaluation team have not made any changes to the objectives of the evaluation from the description and request provided in the UNICEF ToR. The description of the objectives in the report offer a little further detail than the ToRs, explaining that the evaluation will emphasise inclusivity, specifically gender and disability.
	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	The evaluation scope is presented under its own clear subheading, which facilitates navigation for the reader. Again, very helpfully, the scope is divided into subsections, describing geographic and chronological scope. The description of geographic scope provides a rationale for the regions in which the evaluation was conducted. The report states that the sample chosen was an "adequate representative of the GPE supported schools". However, because the context did not include a description of the intervention locations, including how many schools were under the intervention, the reader cannot assess whether the geographical scope was adequate. The chronological scope subsection states that the evaluation covers the entire duration of the intervention period, including the no-cost extension. This is the first mention of a no-cost extension. This could have been included in the earlier description of the intervention.
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	Earlier in the report, when the Object of the Evaluation is presented, Table 1 presents the "focus" of each of the three grants that make up the intervention. The focus can be interpreted to present the expected outcomes and the activities contributing to each of the outcomes. Therefore, this table goes somewhat to describing the interventions intended results. The evaluation team may have considered presenting this information differently. The first suggestion would be to present this information separately, in a section that clearly describes the intervention logic / the intervention theory of change. A further suggestion would be that the outcomes and activities are clearly labelled as such.
	ii Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).	No	There is a small box at the end of Section 5 that offers a very general overview of the intervention theory of change. The description only states that the programme leverages outputs that lead to desired results, and that monitoring will support achievement. Therefore, the description is too general to understand how the intervention was supposed to work. As stated above, Table 1 presents the "focus" of each of the three grants that make up the intervention. The focus can be interpreted to present the expected outcomes and the activities contributing to each of the outcomes. There is no articulation of the expected outputs, leading to the outcomes. As suggested above, the evaluation team may have considered presenting the information in Table 1 differently, as this provides some level of description of how the intervention theory, and its causal relationships. In addition to the suggestions above, the team may have considered presenting a graphic form of how the intervention was hypothesized to work.
	iii For theory-based evaluations, the theory of change or results framework is assessed, and if requested in the ToR, it is reformulated/improved by the evaluators.	No	As mentioned above, there is a small box at the end of Section 5 that offers a very general overview of the intervention theory of change. The description only states that the programme leverages outputs that lead to desired results, and that monitoring will support achievement. Therefore, the description is too general to understand how the intervention was supposed to work. The UNICEF ToR requests that, "the GPE's Theory of Change will guide the evaluation" ... and that the evaluation is expected "to assess the validity of the underlying assumptions and the effectiveness of the strategies employed". The ToR provides a basic "overall GPEP logic" - the theory underlying each the programme and each of the two main components. This theory of change has not been assessed or reformulated by the evaluation team.
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	41%	Comments on Rating
Question 8.	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? <i>UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender and human rights based approach (these can be mainstreamed into other criteria). Humanitarian evaluations should also consider Coverage; Connectedness; Coordination; Protection; Security.</i>		

	i Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Partially	The evaluation draws on the use of all the DAC criteria and has evaluation questions appropriately aligned with each of the criteria. In total, there are 34 evaluation questions. This could be considered too many for such an evaluation; the evaluation team will have stretched the resources and methodology of the evaluation to answer all the questions, and it will have been difficult to answer each question well, whilst delivering a concise, useful report. The number of questions is, however, the number that were posed in the ToR. It is suggested that the evaluation team may have followed the DAC Criteria Principles, considering, for example, "The criteria should not be applied mechanically. Instead, they should be covered according to the needs of the relevant stakeholders and the context of the evaluation" (See OECD DAC Evaluation criteria). In this way, and following the strong accountability purpose of the evaluation, the evaluation team may have prioritised questions aligned with efficiency, effectiveness and relevance, over questions aligned with other criteria.
	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.	No	The evaluation matrix is presented as Annex 1. The matrix lists all the questions and sub questions, and then has a column for indicators, methods of data collection, data sources, and approach to data analysis. This, therefore, forms a strong structure for the matrix. The column for indicators has been filled in / completed for the relevance (sub)questions – all these indicators are quantitative in nature, but the methods of data collection are qualitative. The indicators for the evaluation questions of the other criteria the report references the GPE Results Framework or other indicators, such as "Sample Case Study" or "To a very large extent, large extent, moderate extent, less extent, none". The other columns with methods of data collection, data sources, and approach to data analysis are all completed with exactly the same information. Therefore, looking at the information in the evaluation matrix, it is questionable whether the information in the columns will support a robust 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	The evaluation design is described in two different places. In one place, the evaluation states that it employed a mixed-methods approach, with guiding rights-based framework. In the other place, the report states that the evaluation adopted a post-test intervention design. These two approaches are complementary – although "mixed methods" are methods, rather than a design. Nevertheless, if they were presented in one place this may have provided a more coherent account of the approach. Similarly, the methods described are presented in several different places, with information such as data sources often repeated, rather than providing one coherent account. This notwithstanding, the methods are suitably robust for the purpose and objectives.
	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	The report presents the quantitative data sources in several different places, rather than one coherent place. The report explains that quantitative data is secondary data, sourced from existing databases. The report lists these data sources in the main body of the report. The qualitative data sources are similarly named in several places, rather than one coherent place. All these data sources are suitably, if not repetitively described. However, as up to now in the report, there has been no identification of stakeholders involved in the project, it is hard to assess how appropriate and adequate these sources are.
	iii Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Partially	Quantitative data was sourced from existing, secondary, datasets from implementing partners. It seems, although is not spelled out, that all of the datasets available were used. Therefore, as the report states sampling for quantitative data sources was not necessary. The report explains that the evaluation team drew on both purposeful and convenience sampling strategy for qualitative data sources. The rationale for both strategies is explained and suitable considering the locations for data collection. This sampling approach seemed to support the collection of diverse and informed perspectives. However, the report states that data was collected from 33 donor supported schools. The report does not state the total population of schools supported. Therefore, it is not known whether this sample strategy was representative.

	iv Clear and complete description of the methods of analysis.	Partially	There is a separate section on data analysis. This supports clear navigation for the reader. The description for qualitative data analysis is quite clear, though would benefit from a little more information. For example, the report states that data was analysed thematically. The evaluation team may have considered reporting whether this was 'ground up' thematic analysis or 'top-down' analysis, guided by a coding tree developed from the evaluation questions. There is a good description of the methods of quantitative data analysis – the secondary data collected. The report states that descriptive statistics were used, as well as trend analysis. The report also states that three advanced statistical tests were carried out: comparative analysis, significance testing (T-tests), and gender and disability sensitivity. This description is a bit unclear, as for example, its not clear what statistical test was used for the comparative analysis (was is the significance testing?); and the gender and disability sensitivity analysis is not clear – the data was disaggregated, which is an expected analytical process, but whether there a further analysis is not clear. The description states that for the comparative analysis, non-intervention schools were used – however, were the schools of comparable in terms of context, scale? Do the comparative schools represent a diverse and balanced sample? None of this is clear.
	v Methodology allows for drawing causal connections between outputs and expected outcomes.	No	As described above, the report states that it employed a mixed methods approach and a post-test intervention design. The methodology that is described does not offer a robust test of the theory of change of the intervention nor does it offer a robust pre and post-test. Therefore, the methodology does not allow for drawing causal connections between outputs and outcomes. The evaluation team may have considered employing a strong test of the intervention's theory of change to support an analysis of 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	The evaluation report provides a clear section that describes limitations and the mitigations. Within this section, the evaluation team provide five clear limitations. This is very helpful to the reader and supports understanding. Within the five limitations, there are some strong, valid descriptions of constraints to the evaluation, such as biases in data collection, geographic and security challenges. Whilst the report notes some methodological imitations, the evaluation team may have also considered issues such as: lack of baseline comparison, lack of a robust test of the intervention theory of change.
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	The report has a separate clear section describing ethical issues ad considerations. Within this section, the report describes that the evaluation process complied with UNEG general norms, as well a UNICEF's procedural and ethical standards for evaluation. In addition, the report clearly describes what compliance entailed, following the key principles of UNEG guidelines.
	ii Description of ethical safeguards for participants appropriate for the issues relevant to methodology and how they are applied (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm). For those cases where the evaluation involves interviewing children, explicit reference is made to the UNICEF procedures for Ethical Research Involving Children.	Partially	Within the ethical issues section, the report offers a general description of the ethical safeguards followed. This includes safeguarding and child protection considerations and adherence to UNICEF's Procedures for Ethical Research Involving Children. Although not stated in this section, the evaluation team also obtained approval for data collection from the relevant regional ministry.
Question 11.	Does the evaluation incorporate innovative practice that adds value to the evaluation process?		
	i Innovation practice is used to improve the quality of evaluation process. This could be evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	No	There were no apparent innovative practices to the evaluation process.
SECTION E:	EVALUATION FINDINGS (weight 25%)	50%	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	<p>The relevance findings start by describing the challenges facing the education sector in Puntland. These would have been well placed in the context section of the report – therefore leaving the findings section to describe the intervention's relevance for and effects on the challenges. The findings section presents evaluation questions in a separate colour, separating the findings from each question. The report does not provide findings for every evaluation question put forward in the methodology section. This is unsurprising considering the large number of evaluation questions, however, this does mean that the evaluation does not answer all the evaluation questions. There are also a couple of evaluation questions that are slightly different from those presented in the methodology section. Ideally, for good practice, it is recommended that evaluation team are consistent with the use of evaluation questions throughout the report.</p> <p>Despite these limitations, the findings section is generally well written, with clear comprehensible sentence structure. This makes reading fairly easy and supports the reader learning from the report. Furthermore, the way the evaluation findings are structured is quite helpful. Following presentation of the evaluation question, the report offers a short summary finding, offering an answer to the question. Then, below this, the report provides supporting evidence for the main finding. This way of structuring the findings supports comprehension and reading. The evidence presented is relevant and supports the findings, however, there is little reference to the sources or the strength of the evidence – thus it is unclear whether the evidence comes from data from FGDs, KIIs, or document review. Furthermore, because the evidence sources behind the sentences are unclear, the reader is unsure of whether evidence has been triangulated from multiple sources. The report does not, for example, say how many sources or how many different types of sources evidence a finding.</p> <p>The findings section is generally well written, with clear comprehensible sentence structure.</p>
	ii Explicit use of the intervention's results framework/ToC in the formulation of the findings.	No	<p>The findings section starts by stating that, "This section presents the evaluation findings guided the GPE Theory of Change (ToC)...". However, it is unclear how the ToC or results framework has supported the formulation of the findings. The effectiveness evaluation questions appear to provide findings against the activities and outputs of the intervention – thus leading to evidence for the outcomes, however, activities and outputs are never explicitly referenced. Similarly, the findings do draw on the data within the results framework but generally indicators and targets are not referenced. Therefore, despite well-articulated evidence, it is unclear whether the results are meeting the expectations.</p>
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.	Partially	<p>The evidence presented is relevant and supports the findings; it offers a strong range of qualitative and quantitative data. However, as explained above, there is little (but not no) reference to the sources behind the evidence presented. Thus for the most part, it is unclear whether the evidence comes from data from focus group discussions (FGDs), key informant interviews (KIIs), or document review. However, there seems to be a good range of both qualitative and quantitative evidence provided to support the overall findings. Also as indicated above, although outcome level data is reported, particularly in the effectiveness questions, outputs and output-level data is not explicitly referenced.</p>
	ii Findings are clearly supported by, and respond to, the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Partially	<p>Findings clearly respond to and are a good interpretation of the evidence provided. Although generally reporting positive findings, they do report a balanced perspective. As mentioned above, the findings do draw on the data within the results framework but generally indicators and targets are not referenced. Therefore, despite well-articulated evidence, it is unclear whether the results are meeting the expectations.</p>
	iii Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings also analyse the logical chain (progression -or not- from implementation to results).	Yes	<p>Quite a good range of contextual, organisational, and managerial factors are presented as evidence to explain the findings. There is particular sensitivity to contextual issues, such as COVID-19 and how the pandemic impacted upon the intervention. Similarly, the evidence strongly considers managerial issues very well throughout the findings section.</p>
Question 14.	Does the evaluation assess and use the intervention's Results Based Management elements?		
	i Assessment of the adequacy of the intervention's monitoring system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Partially	<p>There is no evaluation question that directly asks about the intervention monitoring framework. However, the findings indicate, in several places, the suitability of the monitoring system. For example, the report states that the program leveraged the Education Sector Strategic Plan framework for monitoring, suggesting that the framework supported transparency and accountability, with the regular monitoring reviews also supporting timely adjustments to emerging needs. The report suggests, however, that a monitoring framework with greater focus on qualitative assessment, alongside the strong quantitative assessment would have been helpful. The report also highlights that the monitoring framework considered gender and inclusivity. A specific question assessing the intervention monitoring system would have supported a coherent assessment of the framework.</p>
SECTION F:	EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)	63%	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 offer a well written concise conclusion of the findings against the questions of each of the evaluation criteria. The conclusions address each of the evaluation criteria in turn, providing an easy read and sound understanding of the achievements and challenges of the intervention. Suggestions to improve the conclusions, however, would include reflecting on the purpose and objectives of the evaluation, which include "Examine the extent to which GPEP not only met its targeted goals and objectives but also effectively and efficiently addressed the actual needs of the most affected learners". As in the findings section, the conclusions could be strengthened by assessing achievements against targeted goals and objectives. Currently, the conclusions only assesses achievements against one of the initial targets (enrolment of out of school children). A further objective of the evaluation was to evaluate gender, equity, and disability inclusion. The conclusions appropriately reflect on the extent the intervention incorporated GEDI issues into programming, though to further strengthen the conclusions, it would have been good to provide a concluding assessment of "how the programme demonstrated gender responsiveness and addressed the needs of vulnerable groups impacted by emergency". These areas of improvement withstanding, the current conclusions are well written and accessible.
	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 clearly drawn from the findings, whilst providing a high-level assessment against each of the evaluation criteria question areas. There is a well-balanced picture of strengths and challenges of the intervention, which mainly comes from the findings section. Therefore, whilst a strong concluding assessment, there is little insight beyond the findings section.
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	The report offers five strong lessons, which are each preceded by a headline explanation statement. The lessons and explanations are clear and helpful. Following the short lesson statement, the evaluation team have offered a "proposed action". Whilst these are more generic recommendations (than the recommendations that are provided further below in the report), the evaluation team may have considered holding off on providing action points, and adding these to the recommendations section.
	ii Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Yes	As stated above, the lessons are derived from the findings and are a little more generically applicable than the recommendations.
SECTION G:	RECOMMENDATIONS (weight 15%)	75%	Comments on Rating
Question 17.	Are recommendations well grounded in the evaluation?		
	i Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.	Yes	The report presents 19 clearly articulated recommendations that are derived from the findings, but are thoughtful, with good potential for application.
	ii Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Yes	Each of the 19 recommendations appear thoughtful and useful for the stakeholders involved in the intervention. Each recommendation statement is followed by a clear statement offering guidance on implementation, which seem sensible and practical. The main suggestion to improve the recommendations is to signal who is the intended user of each of them. This would have helped uptake of these recommendations.
	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 recommendations are introduced by a short paragraph describing how the recommendations were developed. The report states that the recommendations were developed through a participatory, evidenced based approach, using KIIs and FGDs. It is not clear whether this process was different from the rest of the evaluative process – the data collection period, or whether it was part of the same process. Clarifying this would have supported reader understanding of the process.
Question 18.	Are recommendations clearly presented?		
	i Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.	Partially	The recommendations are appropriately organised into short-term (within 1 year), medium term (1-3 years), and long term (3-5 years and beyond). This classification supports the use of the recommendations. To strengthen the recommendations and support uptake, the evaluation may have considered targeting each of the recommendations to a specific stakeholder.
SECTION H:	REPORT STRUCTURE AND PRESENTATION (weight 5%)	50%	Comments on Rating
Question 19.	Does the evaluation report include all relevant information?		

	i) Opening pages include: Name of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents -including, as relevant, tables, graphs, figures, annexes; list of acronyms/abbreviations, page numbers.	Partially	The opening pages contain all the required information. There are minor formatting issues on the table of contents (e.g. mix of lower and upper case "of" for List of Figure and List of Tables), and the number in the table of contents is missing 6.2. Similarly, there are formatting issues on the List of Tables – which is presented in a different font than the rest of the contents page.
	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 provide helpful, supplementary information for the main report. Suggestions to improve the annexes are to refer to the appropriate annex when relevant in the report (for example, refer to the Annex Terms of Reference when discussing the Terms of Reference). Rather than putting documentary evidence as a bibliography at the end of the report, it may be better to put this as an additional annex.
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).	No	The report is generally very logically structured and easy to navigate. There are clear headings and subheadings; with some sections better structured than others. For example, the Methodology section (Section 6) has 10 subsections. Not all of these subsections are formatted in the same way, some with numbering, some without. This is an issue that runs a bit through the report. Similarly, the subheadings in the Findings sections, under each of the DAC criteria assessed, may have been better presented with numbering the evaluation questions, for easy navigation and reference. Consistency with the use of headings and subheadings and numbering would have helped the presentation of the report. Areas where the structure could be more logically presented include: the methodology section which presents the evaluation approach before the evaluation criteria and questions. This would have benefited from being in different order, as an evaluation approach needs to be developed from the needs of the evaluation criteria and questions. Overall, the report may have benefited from using chapters, with numbered headings and subheadings. For example, the Introduction may have been presented as a chapter, with the national and regional context, and the object of the evaluation as subheadings. Chapter 2 may have been used to present the evaluation, including, as subheadings, the evaluation purpose, objectives and scope. This may have presented a clearer structure for the reader, making it a little easier to navigate between sections. There are also some spacing issues within the report, presenting larger gaps between text and between numbering (e.g. in the recommendations section). There is also some inconsistent indentation of the narrative.
	ii) Structure accords to UNICEF guidelines for evaluation reports: context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.	Partially	The report structure follows the required guidelines in terms of presenting context, purpose, methodology and then findings, conclusion, lessons learnt and then recommendations. It is strongly suggested that preceding the context, there is a short introduction to the whole report, offering a short description of the subject of the report, a couple of lines presenting the intervention, the aim of the evaluation. This sets up the reader with knowledge of why the context is relevant.
Question 21.	Is the report well presented?		
	i) Report is of reasonable length; it does not exceed number of pages that may be specified in ToR.	Yes	The ToR request between 40-60 pages – the main body of the evaluation report is just over 62 pages.
	ii) Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	Generally, the writing is comprehensible. There are some minor syntax, sentence structure, issues that if revisited may strengthen readability and ease comprehension. There are some minor foot note issues. For example, the foot note on page 3 says "ibid", however, there is no preceding footnote on this page, and the footnote on the preceding page does not correspond correctly.
	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 are several tables, and some graphs used in the report. Generally, these are used well to illustrate how intervention metrics have changed over time. Though sometimes both a table and graphic have been presented to illustrate the same statistics. It is suggested that generally either the table or the graph is needed. In other cases, the report has made good use of tables, for example presenting the evaluation questions with criteria and the stakeholder analysis. The report may have also used more graphs to illustrate the large amount of statistical descriptive data that is presented in written form. The report uses one map to locate the intervention within the country. The main map is good, although the insert map, showing Somalia would be stronger if it showed and located Puntland.
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 report states, very clearly, that the evaluation approach was guided by a rights-based approach, aligning with the international benchmarks of the Convention on the Rights of the Child (CRC) and the Core Commitments for Children (CCC). Furthermore, the evaluation team have explained how the principles of the CRC were applied to the evaluation, which is good practice. To strengthen this narrative, the evaluation team may have provided a clearer distinction of applying the rights-based frameworks to the evaluation process, and when it was assessing how the rights-based frameworks were applied within the intervention.
	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	As part of the description of how the evaluation team applied the principles of the CRC and CCC, the report states that the right to participation, were embedded in the evaluation design. The report further states that qualitative data was collected from a diverse range of stakeholders and there was an inclusive approach to understanding the perspectives and experiences of all relevant parties. When describing the evaluation quality assurance process, the evaluation team explained that the inception report was subject to review by the Evaluation Reference Group – which is the first and only time the report mentions a Evaluation Reference Group. Although in the Acknowledgement page, after the title page, the evaluation team acknowledge the support of the group and explain who composed the group. It is suggested that this could have been explained in the main body of the report. The report's description of the level of participation of right holders and duty bearers – indicates that they were included as informants. It is suggested that the evaluation team may have considered more inclusive processes, such as involving rightsholder – particularly children – to be involved with the development of data collection tools; or to involve the reference group in the formulation of recommendations.
	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 used in the report is inclusive, avoiding bias. The language of rights holders and duty bearers is mentioned in the report, but minimally, only in the methodology and recommendations sections. The analysis and findings present a good level of disaggregated data; this could be improved in some sections, but generally groups are fairly well represented.
	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	As mentioned above the intervention is explicit in its intention to assess how the intervention upheld principles of CRC and CCC. The intervention and the evaluation questions help guide the evaluation team in this assessment as the intervention itself and many of the evaluation questions are explicitly about equitable access and inclusion. Furthermore, the methodology section recognises and draws on the LNOB principle – including, for example, stating that the team used experts in sign language to collect data from deaf pupils. Although there is no further reports on the perspective of deaf students, the assessment makes suitable assessment of inclusion, often reflecting disability inclusion.
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 evaluation highlighted that gender considerations were integrated into every phase of GPEP, from identifying priorities and needs to planning, monitoring, implementation, and sustainability efforts. One of the evaluation objectives was to assess gender, equity and disability inclusion, and more specifically to analyse the extent to which the GPEP incorporated gender, equity, and disability inclusion into its design, implementation, monitoring, and reporting processes. There was not a specific evaluation criteria on gender or human rights, but many of the evaluation questions specifically asked about gender equality.
	ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	The evaluation report states that qualitative data was collected from a diverse range of stakeholder involved or impacted by the project. It specifies that data was collected using mixed methods from PWD and marginalised groups and shows that evaluation participants included 42percent females in FGDs and 30percent in key informant interviews (KIIs). The report further explains that data were disaggregated by gender and disability status to evaluate differential impacts and specific indicators compared performance and outcomes for girls versus boys and students with disabilities versus their peers. The report offers a good explanation of how ethical standards were considered throughout the evaluation process, specifically stating that stakeholders were treated with integrity and with respect for confidentiality.
	iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	The evaluation background section is taken from the UNICEF ToR. It provides a short paragraph of intersectional analysis, describing the vulnerabilities of girls, children with disabilities, and children displaced by crises. Findings appropriately included analysed data that has been disaggregated, as well as triangulated with the perspectives of different groups. There are several strong recommendations that are specifically targeted at girls, women teachers, and inclusive practices.

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?