

# GEROS Evaluation Quality Assurance Tool

Version: September 2021

"An integrated and universal social protection linked to Social Protection in South Africa developmental social welfare services in South Africa / short title: Social Protection in South Africa."

## REPORT RATING SUMMARY

<b>Overall Rating</b>	55%	Fair
●●●●● Exceptional (96% - 100%)	5	
●●●●○ Highly Satisfactory (87.5% - 95.99%)	4	
●●●○ Satisfactory (82.5% - 87.49%)	3	
●●○○ Fair (35% - 62.49%)	2	Meets UNICEF/UNEG standards for evaluation reports in some regards, but not all. Decision makers may use the evaluation with caution. Substantive improvements in some areas are needed
●○○○ Unsatisfactory (0% - 34.99%)	1	

## REPORT DETAILS

<b>Title of the evaluation report</b>	"An integrated and universal social protection linked to Social Protection in South Africa developmental social welfare services in South Africa 63-2022-17919"
<b>Report sequence number</b>	South Africa 63-2022-17919
<b>Region</b>	ESAR
<b>Year of report</b>	2023
<b>Office</b>	ILO DWT/Country Office Pretoria and ILO Evaluation Office
<b>Coverage (countries)</b>	South Africa
<b>ToRs present</b>	Yes
<b>Date of review (dd/mmm/yyyy)</b>	8-10 January 2023
<b>Name of review firm</b>	IOD PARC

## CLASSIFICATION OF EVALUATION REPORT

<b>Management of evaluation (Managerial control and oversight of evaluation)</b>	Jointly managed with one or more UN agencies
<b>Unicef goal areas (Alignment with strategic plan priorities)</b>	
Every child survives and thrives	Yes
Every child learns	No
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)	No
<b>Evaluation object</b>	Programme
<b>Evaluation type</b>	Summative
<b>Evaluation strategy</b>	Mixed methods
<b>Evaluation design (primary method used)</b>	Theory-based
<b>Evaluation level</b>	Outcome
<b>Geographic scope</b>	National
<b>Primary SDG(s) covered (number)</b>	1.1, 1.2, 1.3, 5.4, 8.5 and 10.1

**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.

Overall, this report 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. Without addressing some of the major weaknesses of the report, it is unclear whether the findings can help future social protection programming in South Africa.

### Key strengths of the evaluation:

- > A concise but strong contextual overview.
- > Visual aids such as tables and figures are appropriately utilized, clearly presented, labeled, and referenced in the text. This enhances the readability and understanding of key information.
- > Clear analysis of results based on the results matrix under Effectiveness.
- > The report does well in noting causal factors (e.g., COVID, financial constraints, and hiring delays) in the achievement or non-achievement of results which adds context to the findings and helps in understanding the challenges faced by the JP.
- > Annexes provided increase the utility of the report.

### Areas for improvement:

- > A major weakness identified is the lack of supporting evidence for many findings. Several findings are presented without clear data sources or triangulation methods, making it challenging to assess the reliability of the findings themselves and the conclusions that follow.
- > The findings section requires revision to ensure that evidence within a specific finding corresponds directly to the headline finding (i.e., the numbered finding), avoiding instances where information presented under a numbered finding does not align.
- > Additional details are required within the recommendations section to more clearly designate responsible party(ies). As written, 'United Nations' or 'Donors' is too vague to be actionable.
- > The report lacks a clear description of how the recommendations were developed, including the involvement of duty-bearers and rightsholders. Details on the development process, such as feedback from a reference group, would enhance transparency.

**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 topline recommendations are as follows:

### General

- Provide a stakeholder map to illustrate the relationships and reporting lines between implementing partners and government entities, enhancing the understanding of the Joint Programme (JP) structure.
- Include specific information about the contributions and roles of key stakeholders, beyond funding details, both within the UN system and the government.
- Reference changes in objectives from the Terms of Reference (TOR) in the body of the report, explaining the reasons for these additions. This ensures transparency and consistency.
- Provide an assessment of the theory of change, including testing assumptions and a more detailed examination of the underlying logic earlier on in the report instead of within several findings. This will help readers understand the full analysis more easily. This could be built off of the analysis that was supposed to be performed during the inception period of the evaluation.

### Methodology/design

- Reduce the number of evaluation questions to improve focus and feasibility of answering them.
- Clearly explain why a counterfactual approach was chosen, considering resource constraints, and discuss why alternate methods such as contribution analysis were not chosen. Provide more details on how the logically constructed counterfactual was utilized in determining the program's contribution.
- Clearly define and present the sampling strategy used in the evaluation, building on the stakeholder mapping performed during the inception reporting.
- Clearly describe the analytical processes used during data collation and analysis, specifically detailing the methods employed for data computation, tabulation, comparison, analysis, and synthesis as outlined in Figure 3.
- Provide a more comprehensive description of challenges faced during the evaluation and the associated mitigation strategies.
- Clearly articulate how extending the data collection period mitigated specific issues mentioned, such as engagement challenges, hesitancy in providing written answers, and the perception of unfair self-assessments.
- Clearly articulate the role of the evaluation reference group and any quality assurance measures that the report went through, including how recommendations were developed and whether stakeholders took part in their development.

### Findings

- Review and revise findings to ensure that evidence within a specific finding corresponds directly to the headline finding, avoiding instances where information presented under a numbered finding does not align with the headline.
- Clearly identify and describe the data sources used for each finding, including documentation, electronic surveys and interview results. Explicitly discuss how information has been triangulated to ensure credibility and reliability of findings. The report should ensure that findings are complete, supported by evidence, and avoid the inclusion of pseudo-findings that lack sufficient backing. Where evidence is lacking or inconclusive, explicitly state the limitations and refrain from making strong statements without proper support.

Use clear and unambiguous language in findings, avoiding vague terms or statements that may lead to confusion.

– Use clear and unambiguous language in findings, avoiding vague terms or statements that may lead to confusion.

#### Lessons

– Clearly specify which findings contribute to the identified lessons learned, ensuring that each lesson is supported by relevant evidence from the evaluation. As needed, differentiate between lessons derived from findings and those based on the evaluators' general knowledge, providing proper citations and references.

#### Recommendations

– Enhance the clarity of recommendations by explicitly linking each recommendation to specific findings in the report. Provide additional details within each recommendation to offer a more comprehensive understanding of the proposed actions.

– Recommendations should clearly indicate specific stakeholders responsible for implementing each recommendation. As written, the recommendations include very broad entities such as 'donors' and 'United Nations'.

### SECTION RATINGS

SECTION A: EXECUTIVE SUMMARY (weight 5%)		67%	Comments on Rating
<b>Question 1.</b>	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 does a good job of summarizing the programme, purpose and objectives of the evaluation, and the methodological framework of the evaluation. The executive summary should have included only the most key findings – this would have helped cut down on the length of the report. As presented, there are some findings that do not contribute to the conclusions or recommendations and could be omitted from the executive summary. If this approach was taken, a sentence stating that only key or selected findings are presented within the executive summary with more in-depth findings in the sections that follow. To save space, the recommendations could have been presented as a table. Good practices could have been summarized to save space. The lessons learned section refers to box 1, although there is no box 1 in the executive summary. The executive summary contains sufficient information to be a standalone document but exceeds the page limit of 5 pages. The executive summary is 9 pages long.
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 is missing the intended audience.
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.	Yes	The executive summary contains key information required to understand the intervention and evaluation. Based on the quality assurance review, the executive summary does not introduce new information from what is presented in the report.
<b>SECTION B: BACKGROUND (weight 5%)</b>		<b>64%</b>	<b>Comments on Rating</b>
<b>Question 2.</b>	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 report includes the timeline of the programme (which conveys the implementation status), the programme's funding and that it covers the whole country.  That said, there are inconsistencies in the timeline of the programme. The report would benefit from clarifying the implementation period of the programme as there are some references that cause confusion. For example, it is made clear that the programme ran through September 2022 on the title page ('Project duration: January 2020 – September 2022'), in the executive summary and in the introduction. However, the report states under the scope section that the evaluation 'covered the entire programme implementation period from January 2020 to May 2022'. The report should have made clear that the final evaluation was conducted prior to the programme's end and that while the programme continued on until September 2022, the evaluation only covers the period of programme implementation up to May 2022 (if that is the case, as data collection activities may have captured activities that took place after May 2022). Within finding 9, the report reveals crucial information that should be explained earlier on in the evaluation report so that the reader understands the context: 'The SPISA-JP also shifted to Computer Assisted Telephonic Interviews (CATIs) for the fieldwork for national focus group discussions (e.g., for the development of the delays in delivering some of the outputs owing to the COVID-19 pandemic also resulted in a no-cost extension of the SPISA-JP to September 2022.'
ii	Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cities, sub-regions...) and in terms of numbers reached, with disaggregation by gender, age, disability... (as appropriate to the purpose of the evaluation).	Partially	The report contains clear and relevant descriptions of intended rightsholders (beneficiaries) which are titled as 'target groups' in the report. The report does not contain an overview of duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (e.g., institutions/organizations) within the context section. The report lists the JP's implementing agencies in the introduction, but should have also included a list of JP partners.  The report references duty bearers ('ILO, PUNOs, and relevant agencies of the Government of the Republic of South Africa') and rights holders ('intended beneficiaries through their representative organizations, including those representing women and youth in the informal sector') within the section on 'conceptual and methodological frameworks of the SPISA-JP Evaluation'. This information should have been provided in the context section to improve logical flow of information.  The report language indicates that the intervention is a national programme, but the report could have been clearer in this regard. The report did not include geographic locations or number reached. Given the programme's four outcomes and its results framework, numbers reached is not appropriate.
<b>Question 3.</b>	Is the context of the intervention clearly described?		
i	Clear and relevant description of the context of the intervention (i.e. relevant policy, socio-economic, political, cultural, power/privilege, institutional, international factors) and how context relates to the implementation of the intervention.	Yes	The report does a great job at summarising the national context both in terms of larger systemic contextual factors as well as government strategies and UN programmatic contexts.
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Partially	The report clearly states the related SDG targets in the opening pages of the report. The evaluation sought to assess the programme's contribution to the achievement of the SDGs (through SDG targets). The findings include reference to the linkages and contributions of the programme to the SDGs.  Linkages between the SDGs and the programme are clear. However, information related to linkages between specific SDG targets and the programme within the context section of the report is missing. The report would have benefited from clearly detailing the linkages between the SDG targets and the programme within the context section so that the following findings were better framed.
iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	The 'country context' and 'government strategic context' sections provide sufficient detail regarding the situation and needs of the rightsholders/beneficiaries of the JP.
<b>Question 4.</b>	Are key stakeholders, their relationships and contributions clearly identified?		

	<p>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).</p>	Partially	<p>JP implementing partners are stated in the introduction. Other JP partners are not referenced (e.g., specific government entities engaged).</p> <p>As mentioned above, the report does lay out some of the stakeholders within the section on 'conceptual and methodological frameworks of the SPSA-JP Evaluation'. This information should have been provided in the context section to improve logical flow of information.</p> <p>The report references duty bearers (ILO, PUNOs, and relevant agencies of the Government of the Republic of South Africa) and rights holders (intended beneficiaries through their representative organizations, including those representing women and youth in the informal sector) within the conceptual and methodological frameworks section of the report, but does not provide linkages between them in the form of a stakeholder map. It would have been beneficial for the report to provide a stakeholder map to better understand how the JP was structured in terms of reporting lines (e.g., which implementing partners were responsible for specific relationships with certain government entities, etc.).</p>
	<p>ii Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.</p>	Partially	<p>The report lacks specific information about the contributions or roles of key stakeholders, aside from presenting funding details (ILO and UNICEF) and mentioning that ILO was the lead agency. Including additional details about the specific contributions and roles of key stakeholders within both the UN system and the government would have provided readers with a more comprehensive understanding of the context.</p>
<b>SECTION C: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)</b>		50%	Comments on Rating
<b>Question 5. Is the purpose of the evaluation clearly described?</b>			
	<p>i Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.</p>	Yes	<p>The introduction clearly indicates why the evaluation is being performed (following the Guidance Note on Joint Programmes). The 'purpose and objectives of the evaluation' provides sufficient details on the evaluation's intended uses and references how the recommendations and lessons learned will be utilized by the donor, PUNOs and national stakeholders.</p>
<b>Question 6. Are the objectives and scope of the evaluation clear and realistic?</b>			
	<p>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).</p>	Partially	<p>The report clearly and completely describes the objectives of the evaluation.</p> <p>No changes from the TOR were referenced. However, a comparison of the objectives listed in the TOR and the final report did reveal some differences. The final list of objectives include the addition of two objectives:  --Assess the extent to which the SPSA-JP successfully integrated gender equality and human rights as cross-cutting concerns, particularly regarding people with disabilities, throughout its deliverables and process, as well as social dialogue, international labour standards and just environmental transition.  --Assess the impact of COVID-19 on programme implementation, effectiveness, and efficiency.</p> <p>These changes should have been referenced in the body of the report with a short explanation of why/how these were added.</p>
	<p>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).</p>	Partially	<p>The report would benefit from including language around the temporal/chronological scope as it relates to the timeline of the programme (through September 2022) and the timeline included as part of the evaluation (May 2022). The inconsistencies throughout the evaluation report have already been noted above in Question 2.i. The report should include text on why the evaluation only included information up to May 2022 and how this was ensured throughout the data collection process. If data from after May 2022 was used, this should also be made clear.</p> <p>The report is clear regarding what is included thematically and geographically.</p> <p>The reasons for this scope are not explained. As mentioned above, including reasoning for the temporal scope of the evaluation versus the programme's timeframe would have been beneficial. Additionally, including reasoning behind attributing results to the ILO-UNDP-UNICEF-UN Women-OHCHR partnership and programme development partners as a collective would have been beneficial to understanding the evaluation's logic. The scope per the ToR did not make this request, and it is unclear why this approach was taken. It is difficult for the evaluation to fulfil its goal of providing recommendations that are strongly linked to the findings of the evaluation and 'should provide clear guidance to stakeholders on how they can address them' (per the ToR) if findings do not account for different stakeholders' contributions, shortcomings, good practices, etc.</p> <p>Under the scope section, the report states: 'The evaluation also assessed the extent to which the SPSA-JP, through the DaO strategy, leveraged the capacities of PUNOs to mobilize resources for future social protection programming.' However, the reviewer cannot find evidence of where this was done within the report.</p>
<b>Question 7. Is the theory of change, results chain or logic well articulated?</b>			
	<p>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.</p>	Partially	<p>Tables 4 through 7 within the findings section detail the outcomes, outputs and indicators alongside baseline data, targets for October 2022 and status as of May 2022. It would have been beneficial for the report to display all intended results in one table following the theory of change.</p>
	<p>ii Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).</p>	No	<p>The report does not clearly present the causal relationships between outputs and outcomes. The theory of change is presented and a narrative is provided, but these do not clearly match up. It appears that the narrative has attempted to make sense of the theory of change, but the theory of change itself does not clearly state activities/inputs, outputs, outcomes and impact. As the report rightly highlights, the results matrix of the programme lists outputs as outcomes (the report phases this as 'the outcome level indicators were set too low or lacked adequate detail to be truly meaningful to underpin the ToC'). It is unclear why this is the case, and the report does not assess this issue. A revised ToC was required in order to do this analysis. It is unclear why the evaluation did not present the inadequacy of the ToC as a key finding.</p>

iii	For theory-based evaluations, the theory of change or results framework is assessed, and if requested in the ToR, it is reformulated/improved by the evaluators.	Partially	<p>According to the ToR, the inception report was expected to contain an initial analysis of the program's theory of change. However, upon reviewing the inception report, it is evident that no such preliminary analysis of the theory of change was conducted. The ToR did not contain a request to reformulate or improve the ToR.</p> <p>The evaluation report contains two findings related to the theory of change: 1) Under the heading 'evaluability and validity of the design': The finding states that outcome-level indicators were set too low or lacked adequate detail to be truly meaningful to underpin the programme's theory of change. The evidence presented under that finding does not support the headline finding (finding 1) and it is unclear what evidence supports finding 1. The report simply describes the theory of change and its associated assumptions. The report then provides analyses the theory of change's assumptions based on information from 2022. 2) Finding 7: Assesses the progress against the theory of change.</p> <p>The report lacks a comprehensive assessment of the theory of change. While the findings include key components that could contribute to an assessment of a theory of change including testing assumptions (or at least appearing to test the assumptions as updated assumptions were given - but the process of testing assumptions is not explicit within the report) and including some examination of the underlying logic, the report lacks the detail required for a full assessment.</p>
<b>SECTION D: EVALUATION DESIGN AND METHODOLOGY (weight 20%)</b>		55%	Comments on Rating
<p><b>Question 8.</b> 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></p>			
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.	Yes	<p>The evaluation questions are too many in number for the given evaluation with a limited level of effort and timeframe for data collection. Despite this, the evaluation questions listed correspond to the listed evaluation objectives and purpose of the evaluation.</p> <p>The evaluation questions, as presented in Annex 2, correspond to the relevant criteria and these criteria are aligned with the questions. There is some overlap of questions e.g., between relevance and coherence. The inter-relatedness of some evaluation criteria causes this. While this evaluation chose to list only one criterion per evaluation question, another acceptable approach would have been to list the evaluation question then in the next column list the corresponding criteria which could include multiple OECD/DAC criteria.</p>
ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.	Yes	Annex 2 contains the key evaluation questions and sub-questions alongside their indicators, sources of data, data collection method/tools, stakeholders/informants and analysis and assessment. The evaluation matrix does not contain benchmarks nor assumptions. Including the questions, indicators, sources of data, data collection methods and stakeholders for each question are sufficient for this evaluation.
<p><b>Question 9.</b> Does the report specify adequate methods for data collection, analysis, and sampling?</p>			
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	<p>The evaluation design attempted to construct a counterfactual. Given the resource constraints of the evaluation (LOE of about 50 days) and the major challenges in establishing a counterfactual, it is unclear whether establishing a counterfactual was the appropriate method. A contribution analysis may have been a more appropriate methodology to utilize. Conversely, assessing the logical connections within the theory of change could have also helped to infer contribution, even in the absence of a counterfactual. It is unclear why a counterfactual approach was chosen – it was not called for in the ToR and the inception report stated that a counterfactual would 'try to establish what the situation would have been in the absence of the SP-JP, i.e., the counterfactual' but do not detail how this would be done. It is unclear how the evaluation managers could have approved the methodological approach without it being explained in the inception report. The final evaluation report provides little details on the process 'The evaluation utilized a logically constructed counterfactual drawing on process tracing for each part of the ToC and key informant interviews to compare the before and after situations.' The report lists the steps taken to achieve a 'logically constructed counterfactual' but out of the 9 steps listed, only two relate to the counterfactual (8-9). The first seven steps simply relate to a desk review related to the topic under evaluation and an assessment of the ToC and results framework. Upon review of the effectiveness section of the findings, it is unclear how the counterfactual was utilized when determining the contribution of the programme.</p> <p>There are inconsistencies in the description of the evaluation analytical framework. The report states that there are three categories of evaluation criteria. Category 1 included the four OECD/DAC criteria of relevance, efficiency, effectiveness and sustainability. However, the analytical framework visual (figure 2) includes two additional criteria: evaluability and outcomes and impact. Category 2 criteria evaluated the extent to which the programme design and implementation strategy strengthened the strategic, systemic, and national coherence. However, the analytical framework visual presents strategic positioning, systemic coherence and corporate coherence. It lacks national coherence and strategic positioning is a different criterion than strategic coherence. Strategic positioning refers to the unique and advantageous place an organization occupies in its competitive environment or market. Strategic coherence, on the other hand, refers to the alignment and consistency of various elements within an organization's strategy. Category 3 criteria evaluated the degree to which the SPJA-JP mainstreamed cross-cutting issues, including gender equality, a rights-based approach and social dialogue. However, the analytical framework visual lists additional cross-cutting issues including tiparism, international labour standards and just environmental transition. (Note that these topics were not addressed in the findings.)</p> <p>It is unclear why the evaluation states 'The evaluation was conducted as a theory-based programme, focusing on outcome-level indicators and using mixed methods' but then offers data on output-level indicators as of May 2022.</p> <p>Regarding the set of methods used, the report describes desk review, output/outcome/impact indicator measurement tool and primary data collection (key informant and institutional interviews and electronic surveys). These methods are appropriate for an evaluation of this size. The report provides data collection tools as an annex, and the annex should have been referenced within the section describing the data collection methodology. The report does not clearly and fully describe the purpose of the data collection tools or how institutional interviews differed from those held with key informants. Further, the report did not give an overview of the purpose or scope of the electronic surveys.</p>
ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Yes	The report briefly describes the data sources in figure 3 and these appear to be appropriate for his evaluation. For desk review, these included review of programme reports and key strategic documents on social security in South Africa, ILO guidance notes, and relevant strategic documents of PUNOs. For interviews, the evaluation engaged key stakeholders, partners and donors (listed in annex).

	iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	No	A sampling strategy is not provided. The report would have benefited from including the sampling frame by building this out from the stakeholder mapping performed as part of the inception reporting (table 4). The inception report notes that the list would be confirmed and modified during the inception meeting to be held with the Evaluation Management Team (EMT). It is unclear what the final sampling approach or frame was.
	iv	Clear and complete description of the methods of analysis.	No	Figure 3 presents the methodological framework. The framework includes the following information around analysis techniques: "Data computation, tabulation, comparison, analysis and synthesis". It is unclear what methods of analysis were actually employed during the evaluation. The evaluation report would have benefited from clearly describing the analytical processes used during data collation and analysis.
	v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Yes	The evaluation utilized a mixed-methods approach, incorporating various data collection techniques to gather both primary and secondary data. These methods included key informant interviews, institutional interviews, electronic services, and document reviews. This methodology allowed the evaluation to comment on and assess the pathways of the ToC, assess assumptions within the ToC, and populate indicators as of May 2022 (as data allowed). Doing so allowed the evaluation team to draw some causal connections between outputs and outcomes as articulated in the ToC.
	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).	No	While the report contains a limitations section, the section has major deficiencies that required a 'no' rating to this criterion instead of agreeing that the report 'partially' addressed the criterion. Please see detailed feedback below.  The report contains a section titled 'limitations of the evaluation'. This section highlights some challenges but does not provide a clear and complete description of the limitations and associated mitigation measures. For example, the mitigation strategy for the challenge around stakeholders feeling the self-assessment tools were 'unfair' was not given. The only mitigation strategy given in this section is that the evaluation extended the period of data collection through 10 August 2022. It is unclear how extending the data collection period mitigated the issues highlighted such as 'engagement with the inception report and preliminary findings', the 'limited ability of the evaluators to engage stakeholders in discussions on areas where the emerging narratives were contradictory or misaligned', the fact that respondents were hesitant to provide written answers, and the challenge around the perception of 'unfair' self-assessments.  Further, the evaluation could have more clearly stated what was meant by the limitation it calls 'engagement with the inception report and the preliminary findings'. For example, does this mean the limitation was that the client did not give feedback on these two products? There is no further information related to this limitation within the section.  This section concludes that despite these limitations, the evaluators 'are satisfied that adequate information was gathered to adequately answer the evaluation questions'. This statement contradicts evaluation findings which highlight the impacts of data gaps to answering some evaluation questions. For example, the finding that expenditure data was unavailable made it impossible to evaluate efficiency of some outcomes. Additionally, some cross-cutting issues included in the evaluation question matrix were not answered e.g., around international labour standards, just environmental transition and social dialogue.  This section required more information to satisfactorily detail the challenges and mitigation strategies of the evaluation.
<b>Question 10.</b> 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	The report does not contain explicit and contextualized reference to the 2020 revised UNEG Ethical Guidelines within the report itself. However, the report does state that the UNEG ethical guidelines were followed by the evaluation team.
	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	The report contains short descriptions of appropriate ethical safeguards employed during the evaluation (e.g., confidentiality, data protection).
<b>Question 11.</b> 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	The evaluation utilized common data gathering, analysis and dissemination techniques. There was no evidence of innovative practices.
<b>SECTION E: EVALUATION FINDINGS (weight 25%)</b>			<b>58%</b>	<b>Comments on Rating</b>
<b>Question 12.</b> Do the findings clearly address all evaluation objectives and scope?				

	<p>i Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions, sub-questions and criteria.</p>	<p>Partially</p>	<p>The findings address the 11 evaluation objectives, but do not address all of the evaluation's questions and sub-questions.</p> <p>Other feedback on the findings include:</p> <p>The findings were numbered which helped with organization. The headline findings were contained in boxes which is a useful format to help draw the reader's attention to the finding. Tables 4-7 detailing progress against outcome and output indicators were also very helpful in quickly understanding the programme's progress.</p> <p>The findings contain reflections on the ToC that should have been contained in a previous section where the ToC was displayed and explained. For example, the statement that progress for outcome 3 under finding 6.1 could not be assessed because 'it was unclear how the three outputs and their related indicators were intended to logically feed into the outcome indicator. Consequently, results could not be rated from the cumulative performance of the output indicators'.</p> <p>There are some headline findings that contain multiple findings listed in one finding (e.g., two findings listed within the box of finding 8). There are also some instances where the headline finding does not correspond to evidence presented within the body of that finding. In other words, the items discussed following the numbered finding boxes presents separate findings that should be highlighted on their own instead of tucked under the current findings. For example:</p> <p>--Finding 1 is related to the fact that the programme was evaluable with defined outcomes and indicators but that outcome-level indicators were set too low or lacked adequate detail to be truly meaningful to underpin the programme's ToC. The body of the finding then 1) speaks to alignment with the NDS 2030 and SDG achievement as well as some of the SPSA-JP aspirations 2) speaks to the programme assumptions, and comments on the continued relevance of the assumptions.</p> <p>--Finding 3 under the subheading 'stakeholder participation in the programme design' does not relate to finding 3 which specifically addresses the gender and youth dimensions of the programme. Under finding 3, there is also mention of persons with disabilities, but this is not captured in the headline finding that is boxed.</p> <p>--Finding 5 is another example of the headline finding not matching information presented within the section.</p> <p>--Nearly all other findings contain multiple findings within them.</p> <p>There are also some headings which lack a headline finding. For example, the subsection titled 'adequacy of the results and M&amp;E frameworks' between finding 1 and 2 warrants its own finding as it does not relate to finding 1 but it does not correspond to a numbered finding. These are just a few examples of how the findings section is both well structured in terms of its numbering formatting but also poorly structured as it does not number individual findings appropriately. This causes confusion throughout the findings.</p> <p>There are also instances where it is difficult to understand what the report is stating. For example, under the heading 'adequacy of the results and M&amp;E frameworks', the report states: 'Following the Joint SDG Fund guidelines, the SPSA-JP was supposed to serve a two-pronged purpose: (i) Strengthen UNCT capacities for joint programming (operational outcomes); and (ii) Deliver strategic social protection outcomes (programme outcomes). The SPSA-JP showed that the programme was the UN's strategic intervention to align and respond to the government's development priorities in social protection and social security and explicitly link to relevant SDGs. There is, therefore, sufficient clarity regarding its intent. Hence the SPSA-JP was also evaluable on this front.' ---The report then goes on to state: 'The available SPSA-JP annual reports contained relevant data based on its indicators and yearly targets.' It is unclear how the intent of the programme relates to the annual reporting.</p> <p>Another example of unclear language in the findings is under finding 6.3, outcome 3: 'Overall progress on this outcome's performance indicator could not be adequately ascertained because it was not clear how the three outputs and their related indicators were logically intended to feed into the outcome indicator. As a result, the outcome indicator could not be rated from the cumulative performance of the output indicators. The outcome sought to influence the social welfare system for children to strengthen its focus on prevention and early intervention, and the success of the SPSA-JP in this area was limited. Nevertheless, although this result could be attributable directly to the work of the SPSA-JP, the government mandated that social welfare be better funded through the establishment of separate prevention and early intervention.' The link between attributing the (limited) success to the SPSA-JP is unclear.</p> <p>In addition to evidence within a particular numbered finding not relating to the headline finding, the findings also exhibited vague language which was difficult to decipher. For example, finding 1 states: 'The outcome level indicators were set too low or lacked adequate detail to be truly meaningful to underpin the ToC.' It is unclear what is meant by 'set too low' or what is meant by 'truly meaningful to underpin the ToC'. This kind of language requires additional explanation.</p> <p>Finding 6.3, output 3.1 is unclear. It states: 'The evaluation observed that there were two parallel processes (Output 1.1 baseline for 2021 and Output 3.2), both of which were aimed at unpacking the social protection floor, its measurement and costing, and feasibility options for a social assistance scheme (basic income grant) for those aged 18-59 years old and probably for children. The evaluation did not see any evidence that the two processes were coordinated and seeking integration. Hence there was some level of duplication of efforts.' However, Output 1.1 baseline is blank in the results framework.</p>
	<p>ii Explicit use of the intervention's results framework/ToC in the formulation of the findings.</p>	<p>Yes</p>	<p>The results framework was presented within the findings and progress against the results framework was measured to the extent possible.</p>
<p><b>Question 13.</b></p>	<p>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.</p>		

	<p>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.</p>	<p>Partially</p>	<p>A vast majority of findings contain no information on the data sources. It is difficult to understand where findings are coming from and whether the evaluation is using credible forms of qualitative and quantitative data. The methodology notes an electronic survey, but results of this survey are not explicitly discussed in the findings. Similarly, findings from the interviews are also not explicitly discussed. Throughout the findings, it is unclear how information has been triangulated.</p> <p>The findings section contains findings that are incomplete and lack supporting evidence. For example, the finding under the subheading Stakeholder participation in the programme design under finding 3 states that the list provided to the evaluators for sampling participating partners/stakeholders for interviews did not include some of the key institutions. Without presenting any evidence, the report then states that this questioned whether the design intentions of a broad-based participatory SPSA-JP were realized during implementation. This is an example of a pseudo-finding that should have been deleted or edited to reflect a finding based on evidence. For example, if the evidence revealed that key stakeholders such as NEDLAC partners, Provincial Departments of Social Development, South African Human Rights Commission (SAHRC), National Assembly (NA) and the National Council of Provinces (NCOP) and civil society organizations representing labour and people with disabilities were not engaged in the programme at all, then this finding would be stronger. As written, it isn't so much a finding as it is a reflection/question from the evaluation team that does not have sufficient supporting evidence to be a finding.</p> <p>Another example of a finding that lacks sufficient evidence is under finding 5: 'Coherence with development partners' priorities and programmes: The evaluation found that actors in the development sector largely agreed on the need to reform the social protection system in South Africa. There were many separate efforts (including analyses). This had been the case for several years with limited traction on reform. The SPSA-JP appears to have entered the space and stood as an additional voice and effort in a well-trodden terrain.'</p> <p>Another example, also under finding 5 was: 'On probing, some of the respondents acknowledged that there had not been a deliberate focus on the operational aspects of the SPSA-JP. While they indicated the absence of intended measurement and documentation efforts on learning to work together through a joint programme, this did not translate to poor learning on working together. Some anecdotal examples of improved understanding of each other's mandates and opportunities for integrated efforts were cited. A different view was that the evaluation was expecting too much from a team implementing its first joint programme.15 Other expected areas of the SPSA-JP, particularly joint monitoring, and joint communications, were reportedly integrated into similar functions within the PUNOs.'</p> <p>Another example is under finding 6.1 – Statements such as 'The study would form consensus-building engagements' but provides no supporting evidence. Similarly, the same finding states that 'this debate was likely to be sustained even after the completion of the SPSA-JP until consensus on viable options had been built and implemented' but no supporting evidence on why/how the evaluation team arrived at the finding that the debate would be sustained (e.g., plans for government funding of the BIG and BIS initiatives, government commitments to these initiatives, views from government stakeholders and others).</p> <p>Another example is finding 12 which does not provide supporting evidence for the information and statements/opinions expressed.</p> <p>Another example is finding 17 which states 'The evaluation established that the SPSA-JP had a Strategic Communications Plan, but almost all the interviewed PUNO technical focal persons did not know about its existence. Hence the Communication Plan was not consciously implemented' but provides no supporting evidence.</p>
	<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.</p>	<p>Partially</p>	<p>As noted above, there are many instances where findings are not clearly supported by evidence. When the report does include evidence, the authors include both positive and negative findings. Some findings are based on clear performance indicators (e.g., those discussed with the corresponding tables 4-7). The findings often use language that makes it difficult to understand the source(s) of the underlying evidence that support a finding. An example of this is the following statement which does not indicate where the finding was drawn from (e.g., desk review documentation, key informant (singular), multiple key informants?): 'The evaluation noted that as part of the proposal preparation process, CODI was raised as a potential starting point for national engagement on social protection. The programme document does not reference CODI or any other social protection assessment.' There are also some findings that do not appear to be appropriate under the heading. For example, under 'systemic coherence/alignment' under finding 5, the report states: 'There were mixed views on whether the SPSA-JP should have been placed within a broader UN system social protection work portfolio.' And ' Further scope for functional alignment lay in using the Joint SDG Fund ToR...' It is unclear why these are being discussed under coherence and a better framing of the links between these would have clarified this. Under finding 12, there is a list of key enablers in the implementation of the SPSA-JP. That list includes '(b). COVID-19 and the urgent need to find an inclusive response created a policy window for social protection dialogue and reform.' This appears to directly contradict finding 9. Findings would benefit from more clearly laying out the changes the programme made for COVID and discussing these up front in the report. As written, it is during the findings that the report references terms such as 'repurposed programme' and 'toned-down implementation targets' without defining what changes were made. Lastly, the emerging findings listed under the findings section are unclear. Some share information that is not emerging (like the lack of social protections for those aged 18-59). It also contains general questions and statements that don't clearly relate to the joint programme. It is unclear what purpose this section serves.</p>
	<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).</p>	<p>Partially</p>	<p>The report does a good job at noting the causal factors (e.g., COVID, financial constraints, hiring delays, etc.) in achievement or non-achievement of results. The findings include discussion around the logical chain of certain outcomes and how these links or missing links impact the evaluation's ability to assess certain aspects of the programme. The findings section contains many findings that do not have supporting evidence but that make strong statements such as the following: 'Through the decent employment effects of this incentive to informal sector work, the SPSA-JP also contributed to SDG target 8.5. Whilst the results had not yet translated into actual benefits for the target group, the likelihood of success was high.' It is unclear what factors cause the likelihood of success to be high.</p>
<p><b>Question 14.</b></p>	<p>Does the evaluation assess and use the intervention's Results Based Management elements?</p>		

	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>Finding 18 addresses findings related to the M&amp;E system. The evaluation report comments on the lack of a monitoring and evaluation system for the programme. Further, the report found that the SPSA-JP document did not define the coordination, implementation and monitoring roles and responsibilities of the various key governance structures, including RCO, Steering Committee, convening/lead agency and PUNOs. The report found that the SPSA-JP did not have an integrated M&amp;E system, but the system was decentralized to the PUNOs in charge of the various outcome components.</p> <p>In terms of assessing the adequacy of the interventions monitoring system to support decision making, the report offered the following conclusion which would benefit from greater detail and reflection to adequately reflect on the adequacy of the system or the implications of the M&amp;E-related findings: 'Monitoring and evaluation: No integrated M&amp;E system was in place. The M&amp;E function was decentralized to the PUNOs. Because of this, the evaluation could not identify where programme risks and assumptions, cross-cutting issues, and communication and learning plans were being monitored. There was no evidence that the M&amp;E system collected and used data disaggregated by gender, people with disabilities, and age.'</p>
<b>SECTION F: EVALUATION CONCLUSIONS &amp; LESSONS LEARNED (weight 10%)</b>		<b>75%</b>	<b>Comments on Rating</b>
<b>Question 15.</b> Do the conclusions clearly present an objective overall assessment of the intervention?			
	i Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).	Yes	The conclusions are clearly formulated by subheadings and reflect the purpose and objectives of the evaluation.
	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	<p>The conclusions largely appear to be derived from the findings. However, the conclusion 'the utilization of results' includes information that was not presented under Effectiveness Outcome 1 of the report. The conclusions reveal that the 'DSD specifically mentioned the work under Outcome 1 as having contributed to their efforts to push the social protection agenda in national policy forums.' The reviewer cannot locate where this finding was contained within the findings section of the report.</p> <p>The conclusions contain five transformations based on the desk review and it is unclear how these are appropriate for the conclusions section. It is unclear how these present a picture of the strengths and limitations of the intervention. Many of the conclusions simply summarize the findings. However, some conclusions (e.g., coherence) provide additional insight and analysis beyond the findings.</p>
<b>Question 16.</b> 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	While some of the lessons clearly stem from the findings, others appear to be based on the evaluators' knowledge of social protection more generally, and no supporting evidence (e.g., from certain findings or from cited sources) is provided. For example, it is unclear what findings relate to the operational lessons.
	ii Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Yes	The lessons presented are concise but contain sufficient detail to be useful.
<b>SECTION G: RECOMMENDATIONS (weight 15%)</b>		<b>25%</b>	<b>Comments on Rating</b>
<b>Question 17.</b> 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	<p>The recommendations largely align with the evaluation purpose, but all would have benefited from more details. Overall, the recommendations logically derive from the findings. The findings would be strengthened if text was added linking each recommendation to findings (e.g., simply stating below the recommendation 'links to findings x, y and z'). One example of a recommendation that does not clearly link to the findings is recommendation 5 and 6. Recommendation 3 would benefit from more clearly linking it to the programme under evaluation.</p> <p>Recommendation 4 appears to be dependent on the findings from the study in Recommendation 1. This should have been made more explicit. It is unclear how the evaluation is able to make a recommendation following a recommendation to investigate whether and SPIAC-B would be relevant and feasible.</p> <p>More information is required for Recommendation 5 in order to detail how the UN Social Protection Group would differ from the SPIAC-B.</p> <p>Lastly, some of the lessons learned in the annex contain what appear to be recommendations. However, these recommendations are not detailed in the recommendations section. See, e.g., lesson learned 4 which contains this precondition: "Staff should be trained on joint programming for effectiveness and efficiency in the running of the joint programme" and lesson learned 3 which included this precondition: "Need to understand scope of the JP TORs and define a programme scope that adequately achieves intended results within the predetermined timeframe."</p>
	ii Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	No	<p>The recommendations are addressed to entities such as 'United Nations' and 'Donors'. Without clearer guidance on who is responsible, these recommendations are not actionable. No guidance is given for implementation.</p> <p>The report should more explicitly state who is responsible for the recommendation and include some high-level guidance on implementation.</p>
	iii Process for developing the recommendations is described, and includes the Involvement of duty-bearers, as well as rights holders when feasible (or explanation given for why they were not involved).	No	The report does not contain a description of how the recommendations were developed. The involvement of duty-bearers and rights holders is unknown. The report should have contained a summary of the development process (e.g., whether a reference group reviewed and gave feedback, etc.).
<b>Question 18.</b> 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 clearly presented with tables detailing to whom the recommendation is addressed, its priority, its timeframe and the level of resources required. However, many of the recommendations are addressed to the United Nations generally and are not specific enough to support use. The report requires more clearly stating the specific stakeholders responsible for carrying out the recommendation.
<b>SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)</b>		<b>57%</b>	<b>Comments on Rating</b>
<b>Question 19.</b> 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 report contains most of the required information for this criterion. However, it is missing the timeframe of the evaluation.
	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.	Yes	The annexes include the terms of reference, evaluation matrix (titled 'Key SP-JP Evaluation Questions by Evaluation Criteria'), list of interviewees, data collection instruments, and a list of documents reviewed. The report also contains annexes for good practices and lessons learned. However, the structure for the lessons learned annex is inconsistent and does not contain sufficient detail for some of the lessons. The lessons learned annex uses unclear terminology, e.g., 'wicked problems' (See lesson learned 2: "Investment in the process is important, especially where solutions are contested (wicked problems)"). The utility of the lessons learned annex is unclear as it does not contain sufficient detail for each lesson. The theory of change was included in the report body.
<b>Question 20.</b> 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 could have been improved by employing clearer formatting for headings and subheadings to facilitate easier navigation. This could involve using distinct font sizes/colors or implementing a numbering system. Notably, certain subheadings, such as the evaluation criteria 'Orientation to impact and sustainability,' are not included in the table of contents. This omission creates difficulty in navigating to this section compared to other parts of the report. Further, the annexes are not named consistently (see Annex 3, for example).
	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 largely follows what is described within this criterion. However, both good practices and lessons learned follow the recommendations. To enhance clarity, it would be more effective if the report presented good practices and lessons learned before the recommendations. This sequence better aligns with the idea that both good practices and lessons learned contribute to shaping the recommendations.
<b>Question 21.</b> Is the report well presented?				
	i	Report is of reasonable length: it does not exceed number of pages that may be specified in ToR.	No	The report is 50 pages long excluding annexes and the executive summary. The report is concisely written and includes sufficient detail. However, this is in excess of the limit set in the TOR: "A concise final Evaluation Report (maximum 35-40 pages plus 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. The report is generally free from grammar, spelling and punctuation errors, although the reviewer noted some missing punctuation and use of acronyms before they were defined.
	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.	Yes	The report properly utilises visual aids such as tables and figures to help detail key information (10 tables and 3 figures). These are clearly presented, labeled and referenced in the text. The lessons learned section contains some text that is boxed in, however, the box is not labelled. The first paragraph of the lessons learned section refers to Box 1, but the box itself is missing a title. The title should be added in the body of the report and included in the table of contents.
<b>SECTION I: EVALUATION PRINCIPLES (weight 10%)</b>			<b>60%</b>	<b>Comments on Rating</b>
<b>Question 22.</b> 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.	Partially	The report does not reference CRC, CCC, CEDAW or other rights-related benchmarks in the design of the evaluation. However, the report states that the evaluation pursued a human rights-based approach through consultations with duty bearers, development and implementation partners and rights holders.
	ii	Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).	Partially	The report references an evaluation reference group, but details are lacking. The report makes clear that there were consultations with duty bearers, development and implementation partners and rights holders. The report would have benefitted from a clearer description of the level of participation of key rights holders and duty bearers during the evaluation.
	iii	Language is empowering and inclusive, avoiding gender, heterosexual, age, cultural and religious bias, among others; use terminology of rights holders and duty bearers; data is disaggregated by marginalized group; differential results are assessed (distribution of results across different groups).	Partially	The report largely used empowering language with the exception of still using beneficiaries alongside duty bearers in some instances.
	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	The evaluation provided assessment of the extent to which the programme addressed child rights and leave no one behind. The evaluation contained disability-inclusive questions.
<b>Question 23.</b> Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards with detail provided below)				
	i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Fully integrated	a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results? YES b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? YES c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? YES d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? YES
	ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Partially integrated	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? NO b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations? YES c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? PARTIALLY d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? UNKNOWN e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? YES

iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	<p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality? PARTIALLY</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable? NO</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? NO</p> <p>d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area? NO</p>
-----	---	----------------------	--

**SWAP Rating Guidance**

<p><b>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.</b></p> <p>a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?</p> <p>b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?</p> <p>c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?</p> <p>d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?</p>
<p><b>ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.</b></p> <p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?</p> <p>d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?</p>
<p><b>iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.</b></p> <p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described?</p> <p>d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?</p>