GEROS Evaluation Quality Assurance Tool

Version: September 2021

Formative Evaluation of the National Stunting Reduction Movement in Indonesia (2018-2024)

REPORT RATING SUMMARY		
Overall Rating	72%	Satisfactory
● ● ● ● ● Exceptional (96% - 100%)	5	
● ● ● ● Highly Satisfactory (87.5% - 95.99%)	4	
● ● ● - Satisfactory (62.5% - 87.49%)	**3**	**Meets UNICEF/UNEG standards for evaluation reports. Decision makers may use the evaluation with confidence**
● ● - - Fair (35% - 62.49%)	2	
● - - - Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
Title of the evaluation report	Formative Evaluation of the National Stunting Reduction Movement in Indonesia (2018-2024)
Report sequence number	Indonesia/60/2024/18977
Region	EAPR
Year of report	2024
Office	Indonesia Country Office
Coverage (countries)	Indonesia
ToRs present	Yes
Date of review (dd/mmm/yyyy)	February 10, 2024
Name of review firm	IOD PARC

CLASSIFICATION OF EVALUATION REPORT	
Management of evaluation *(Managerial control and oversight of evaluation decisions)*	Jointly managed with Country
Unicef goal areas *(Alignment with strategic plan priorities)*	
Every child survives and thrives	Yes
Every child learns	No
Every child is protected from violence and exploitation	No
Every child lives in a safe and clean environment	No
Every child has an equitable chance in life	Yes
Gender equality (cross-cutting)	Yes
Humanitarian action (cross-cutting)	No
Evaluation object	Strategy
Evaluation type	Formative
Evaluation strategy	Mixed methods
Evaluation design *(primary method used)*	Theory-based
Evaluation level	Output & Outcome
Geographic scope	National
Primary SDG(s) covered *(number)*	2

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.*

Strengths:
--The report is logically structured, well written and is accessible to a wide audience.
--The evaluation report provides a detailed description of the national strategy/object of the evaluation (nearly 20 pages).
--The evaluation report clearly states the evaluation purpose, scope, objectives, 'why now', intended use and intended users.
--The evaluation report provides visual representations of the different iterations of the ToC, including the original and the further revised ToC created during the evaluation.
--The report gathered data from nearly 1,000 respondents from various stakeholder groups, giving the findings a rich data base.
--The findings systematically answer each evaluation question. Where data availability was an issue, the authors clearly indicate this limitation.

Weaknesses:
--The executive summary does not contain sufficient information to be a standalone document useful for decision making.
--The report's length (over 100 pages exclusive of executive summary and annexes). This is not aligned with the ToR which stated: 'In general, the final report should not exceed 70 pages and should aim for conciseness, readability, and visual appeal.' The length of the report may impede uptake of information and reduce accessibility to relevant audiences.
--Lessons learnt/good practices do not contain sufficient detail to be useful.
--The conclusions section provides summaries of findings and does not analyse and interpret the significance of those findings (i.e., does not add insight and analysis beyond the findings which is required of conclusions).
--The recommendations are many (over 60) and would have benefited from a revision to 1) more clearly indicate the order of recommended actions as many appear to rely on another being completed first or alongside another both within and across pillars, 2) remove duplicative recommended actions as there are many actions across pillars that could be completed more efficiently in one exercise, 3) ensure that sufficient details are available to operationalise the recommended action, 4) ensure that the responsible party/ies is/are provided for each recommended action. Reorganising the recommendations to ensure that the order of recommended actions is clear and removing duplicative recommendations would have shortened the recommendations section and created a clearer guidance model for the intended users. As written, the recommendations section does not provide a clear roadmap for execution and the sheer number of recommended actions are likely to overwhelm the responsible party(ies).
--Data collection tools and accompanying informed consent processes are not included as annexes.

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 report would benefit from a thorough copyedit/proofread to ensure that references to tables/figures are correct, table/figure numbering is accurate, annex numbering is accurate, and the header system of the report is properly reflected in the table of contents. The report could consider numbering sections and subsections. As written, there are many subsections and subsections within those that it becomes difficult for the reader to follow.
--The report would benefit from a more detailed executive summary. As written, the executive summary does not contain sufficient detail to be a standalone document used to inform decision making.
--The report would benefit from some revision to cut down on its length. It is currently over 100 pages, exclusive of the executive summary and annexes. This is likely a barrier to uptake/consumption of the evaluation output, and is not aligned with the ToR which stated: 'In general, the final report should not exceed 70 pages and should aim for conciseness, readability, and visual appeal.'
--The report would benefit from providing disaggregated data from the online perception survey (e.g., by stakeholder type) and providing additional disaggregation for evaluation findings based on KIIs and FGDs (e.g., by stakeholder type). Currently, data from the survey and interviews are referred to as one group.
--The authors should include more details on the lessons learned/good practices. As written, they do not contain sufficient detail to be useful.
--The conclusions section should include interpretation of the significance of evaluation findings, adding insights and analysis beyond the findings themselves.
--The recommendations section would have benefited from a rewrite to clearly indicate which activities must take place in which order, reduce duplication of activities and to decrease the number of recommendations which may overwhelm the responsible parties (current number of recommendations exceeds 65). As written, the recommendations section does not provide a clear roadmap for execution and the sheer number of recommended actions are likely to overwhelm the responsible party(ies). (See detailed feedback below.)
--Data collection tools and accompanying informed consent processes should be included as annexes.

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 organized around the five main pillars of the national strategy as well as gender, equity and human rights and sustainability. Each section touches on some of the OECD DAC criteria per pillar/section, but not all criteria. The executive summary provides a very high-level overview of the evaluation findings and recommendations. The executive summary is missing key details required to inform decision making (e.g., responsible parties for each recommendation). Further, there are statements provided in the executive summary that do not contain enough information to be in a standalone document. For example: 'The funding mechanisms for nutrition and stunting were complex and bureaucratic, leading to delays in allocation. Although recent budget-tagging efforts were a step forward, too many budget codes were assigned to non-relevant activities.'; it is unclear what 'too many budget codes were assigned to non-relevant activities' means. This approach is good for a document that is meant to be an evaluation brief (which the ToR states is a separate deliverable from the executive summary) targeted to a wide audience and primarily used for advocacy. However, as stated above, this format does not allow for the executive summary to act as a standalone document useful for informing decision making.
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 includes a brief overview of the intervention, evaluation purpose and objectives, evaluation methodology, findings and recommendations. The executive summary does not include intended audiences, lessons learned and conclusions.
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 noted above, there are some statements in the findings section that lack contextual information to act as standalone findings in the executive summary. More detailed summaries of the findings is required to understand the evaluation findings. Further, details around the responsible party(ies) for each recommendation are not provided. The executive summary does not introduce new information form what is presented in the rest of the report.
SECTION B:	**BACKGROUND (weight 5%)**	93%	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.	Yes	The report fully meets this criterion. It provides a detailed description of the national strategy (object of the evaluation) including locations (all 514 districts in Indonesia), timeline (signed in 2018 and implemented until 2024), budget (table 1), and implementation status (background section).
ii	Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (i.e., institutions/organizations; communities; individuals…), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cites, sub-regions…) and in terms of numbers reached, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).	Yes	The report provides a clear and relevant description of the target population for the strategy including rightsholders and implementing partners by type. Geographic locations are not provided but the background section makes clear that all 514 districts of Indonesia had become priorities for stunting reduction. In terms of national programme coverage and disaggregation by gender, age and disability, the evaluation team was unable to find relevant information on the coverage and breakdown of these.
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 provides a brief overview of the country context as it relates to stunting efforts, including providing statistics on poverty rate, malnutrition and national/subnational government efforts to reduce stunting. The background section also describes the funding context in which the intervention is implemented. The context section would have been strengthened if the authors explored the social determinants of stunting in Indonesia (e.g., maternal education, parental occupation, income, etc.). The context section currently lacks an analysis of sociocultural mechanisms that impact stunting. Further details on the locations where stunting is most prevalent, the number of stunted children under 5 (in addition to the national 21.6% figure overall), and other details to paint a more in-depth picture of the national and local contexts would have also strengthened the context section. Figure 11 notes the underlying determinants of stunting in Indonesia, but this information would have been good to highlight in the context section so that the reader understands the national context more fully prior to reading about specific programme interventions under the strategy.
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Yes	Linkages are drawn to SDG 2 and its target to end stunting.
iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	The report shares the 2022 stunting rate of under-5 children (21.6%), but does not provide data on the location of these children, the population of this under-5 group, etc. While the report meets this criterion by providing a clear and relevant description of the status/needs of the rightsholders/beneficiaries of the intervention, the report would have been stronger with more details regarding children at risk of stunting in the country.
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).	Yes	The report identifies the implementing partners (pg 10) as well as other stakeholders. The coordination structure of the national strategy is provided in figure 1 and figure 2. The roles and responsibilities of different actors in implementing the strategy are available in figure 8.
ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Yes	The budget description (pg 9) includes details on different funding streams (e.g., World Bank). The report does not state that UNICEF provided any specific financial contributions in the background section. Roles and contributions of key stakeholders are described by pillar on pp 10-27. This includes UNICEF's role in supporting the Global Action Plan on Child Wasting and its operational roadmap as well as its status as a development partner for SUN.
SECTION C:	**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	100%	Comments on Rating
Question 5.	Is the purpose of the evaluation clearly described?		
i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	The report clearly describes the evaluation purpose, objectives and scope, including why it was performed at that point in time (i.e., to inform the next phase of the national strategy). The report also details the evaluation users on pp 27-28.
Question 6.	Are the objectives and scope of the evaluation clear and realistic?		
i	Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR (if applicable).	Yes	The report notes the evaluation objectives provided in the ToR and does not reference any changes made to the objectives of the ToR.

ii	Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	The report clearly lays out the scope of the evaluation. Thematic: National strategy/program to accelerate stunting reduction movement. Temporal scope: November 2018-April 2023. Geographic scope: 3 provinces and two districts/cities per province: (1) Central Java province – Grobogan and Wonosobo districts; 2) Bali province – Gianyar and Buleleng districts; and 3) West Sulawesi province – Mamuju and Majene districts). The report provides reasons for this limited geographic scope: limited funds for the evaluation.
Question 7.	Is the theory of change, results chain or logic well articulated?		
i	Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Yes	The theory of change (original) is available in figure 12. The evaluation team assessed these pathways and provided a revised theory of change as part of the evaluation (see figure 30).
ii	Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).	Yes	The original theory of change for the strategy is presented in figure 12. The revised theory of change created by the evaluation team is shown in figure 30. Supplementing this is figure 11 which details the underlying determinants of stunting in Indonesia. The evaluation team provides a narrative form of the theory of change on page 32.
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.	Yes	The theory of change was assessed and revised by the evaluation team. The evaluation team also created a new logical results framework based on evaluation findings to better explain and operationalise the revised theory of change.
SECTION D:	**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	77%	Comments on Rating
Question 8.	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? *UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender and human rights based approach (these can be mainstreamed into other criteria). Humanitarian evaluations should also consider Coverage; Connectedness; Coordination; Protection; Security.*		
i	Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	The overarching evaluation questions per OECD DAC criteria are provided on pp. 31-32. Annex 1 provides the full list of evaluation questions and sub questions. These questions are appropriate for meeting the purpose and objectives of the evaluation. Each question aligns to specific OECD DAC criteria in Annex 1.
ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.	Yes	The final report references the evaluation matrix and provides the matrix in Annex 2. This matrix includes evaluation questions/sub questions, indicators, data collection methods, and data analysis methods.
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.	Yes	The mixed-methods design of the evaluation is relevant and adequately robust for the evaluation's purpose, objectives and scope. The evaluation methods are fully and clearly described on pages 31-38.
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 evaluation used appropriate data sources including secondary data as well as primary data sources including national and subnational actors as well as women of reproductive age. These data sources are a mix of primary and secondary data from both qualitative and quantitative sources.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Partially	The sampling strategy (purpose and random sampling) is described on pp 36-37. The sampling strategy includes a description of how diverse perspectives in terms of a range of implementing partners/key stakeholders were captured. However, the report lacks details on how or whether the evaluation made a concerted effort to capture perspectives of individuals in all of their diversity (e.g., accommodations for persons with disabilities, purposive sampling to include ethnic or religious minorities, persons with disabilities or caretakers of children with disabilities, etc.). The report would have been strengthened by providing details on whether or how diverse perspectives beyond the stakeholder type were captured.
iv	Clear and complete description of the methods of analysis.	Partially	The body of the report lacks a description of the methods of analysis beyond stating that data analysis and triangulation were performed. Annex 2 (evaluation matrix) references the type of data analysis that was planned to take place during the evaluation per evaluation question (i.e., thematic analysis). The final report should have included a summary of data analysis techniques actually employed during the evaluation in the body of the report.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Yes	The evaluation used a mixed-methods approach with both random and purposive sampling techniques. This approach allows the evaluation team to verify or complement findings through triangulation, thus enhancing the evaluation team's ability to draw reliable causal connections between outputs and expected outcomes.
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).	Yes	Pages 37-38 lay out the limitations and constraints faced by the evaluation. Where possible, the report also describes mitigation efforts.
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 includes explicit reference to UNEG ethical guidelines (pp. 34 and 39). The report notes that ethical clearance and research clearance were granted.

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	The report provides a brief reference to the ethical safeguards for participants including ensuring respect for participants such as confidentiality, anonymity and privacy, following the 'do no harm' principle and respecting participants' dignity, diversity and rights through the use of approved instruments, ethics protocol and consent forms. Beyond these brief references, report does not contain descriptions of the ethical safeguards for participants. The report states that interviews were held with 'women of productive age' which could have included those between 15 and 49. However, the report does not explicitly state whether individuals under 18 were included in the evaluation. Additional details regarding the ethical protocols would have strengthened the report.
Question 11.	Does the evaluation incorporate innovative practice that adds value to the evaluation process?		
i	Innovation practice is used to improve the quality of evaluation process. This could evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	No	The evaluation utilised appropriate practices for data collection (e.g., focus groups, interviews, desk review, survey), but there were no innovative practices used to improve the quality of the evaluation process.
SECTION E:	**EVALUATION FINDINGS (weight 25%)**	92%	Comments on Rating
Question 12.	Do the findings clearly address all evaluation objectives and scope?		
i	Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions, sub-questions and criteria.	Yes	The evaluation findings section is quite extensive - at 60 pages in length - and systematically goes through each evaluation question with subquestions as subheadings within sections. The findings are organised by criteria (relevance, coherence, effectiveness, efficiency, sustainability, and gender, equity, and human rights. (Note, however, that the table of contents is not formatted correctly and has sustainability fall under efficiency and is missing gender, equity, and human rights as a header). The findings often provide sufficient levels of evidence to answer each question. However, there are times when evidence is simply unavailable - to no fault of the evaluation team. These instances are noted and the evaluation team is clear about limitations due to lack of available data. Given this, the report fully meets this criterion.
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Yes	In addition to other headers such as gender, equity and human rights, the findings are organised by pillar which corresponds to pillars presented in the intervention's results framework. Strategic outcomes and intermediate results are also discussed within the findings.
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	The evaluation uses credible forms of qualitative data (through desk review and primary data collection via FGDs and KIIs with several stakeholder groups/types) as well as quantitative data (through desk review of government reporting as well as through the perception survey administered by the evaluation team). The evaluation presents output and outcome data per the national stunting reduction strategy evaluation indicators (results framework performance indicators) in the body of the report as well as Annex 4. The evaluation also investigated progress against other indicators relevant to the strategy and presented as annex 5. Triangulation is evident throughout the findings section. The authors consistently use data from stakeholder interviews with desk review data to validate findings. The report uses language like 'many respondents' or 'most stakeholders' without delving into whether these were government stakeholders, community members, etc. Additional details on the types of stakeholders who expressed similar opinions would have further strengthened the findings. The online survey targeted national, provincial and district level line ministry focal persons involved in the implementation of the programme, as well as SUN network members (civil society, business, academia and associations, UN agencies and donors). However, survey results were not disaggregated by stakeholder type which is a missed opportunity for triangulation between opinions of different stakeholder groups.
ii	Findings are clearly supported by, and respond to, the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	Findings present both positive and negative information obtained from document review and primary data collection. These findings are supported by the evidence presented. As appropriate, the findings are based on indicators available in annexes 4 and 5 and detailed in the results framework. The evaluation report systematically assesses each intermediate outcome relevant to the evaluation question. The report also highlights that 'for the next strategic phase, starting in 2024, it will be important to have more SMART (specific, measurable, achievable, relevant and time-bound) indicators in a clearer results framework, as some of the indicators were vague and not clearly measurable.' Annex 4.2 includes progress against targets for the impact, intermediate results and outcome performance indicators from the original results framework.

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	Where data allows, the findings assess factors that led to the achievement/partial achievement/non-achievement of the results (e.g., duplication of efforts, uncoordinated efforts, lack of clearly defined roles/responsibilities, changing guidance on reporting, unclear monitoring/reporting systems at the provincial/district levels, issues with server capacity and infrastructure for online reporting tools/systems, etc.). The evaluation team is also clear that there were some instances where 'locating the data or verifying that the indicator had been met was a challenge. In addition, for some performance indicators, measures were not clearly defined. The evaluation team therefore interpreted the indicator achievement considering what could be verified in terms of documentation.' This is a theory-based evaluation and findings also include an analysis of the linkages between the logic displayed in the theory of change, including how the strategy has progressed through to results. The evaluation team offers a revised theory of change as part of the recommendations section.
Question 14.	Does the evaluation assess and use the intervention's Results Based Management elements?		
i	Assessment of the adequacy of the intervention's monitoring system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Yes	The evaluation report assesses progress of the intervention's monitoring system through intermediate results 5.1-5.4 of the national strategy's results framework (shown in figure 13). The authors include an assessment of the adequacy of this system and notes challenges due to uncoordinated efforts, changing guidance on reporting, unclear monitoring/reporting systems at the provincial/district levels, issues with server capacity and infrastructure for online reporting tools/systems, etc. As part of the evaluation, the team provided an updated results framework based on evaluation findings to better explain and operationalise the revised theory of change with the goal of supporting decision-making. Recommendations related to improvements for the monitoring and evaluation system were also included in the evaluation report.
SECTION F:	**EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)**	25%	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).	No	The conclusions section does not provide conclusions but instead provides summaries of the findings. As such, it does not fulfil this criterion. Detailed feedback on the text presented in the conclusions section follows under question 15.ii.
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 section provides summaries of the evidence (which includes presenting a picture of the strengths/limitations) but does not analyse and interpret the significance of those findings. As such, the evaluation report does not fully meet this criterion as it does not add insight and analysis beyond the findings. An example of where the section nearly provided a conclusion is under financial resources where the authors state: 'Funding mechanisms at the sub-national level were not well understood by relevant stakeholders and did not always support appropriate allocations for nutrition activities or local ownership. Sub-national budget tracking is essential, but it was not operational. It could provide data that all stakeholders could use to hold government to account on their commitments and funding targets.' While the quoted text highlights key problems, it stops short of providing deeper interpretation or making evaluative judgments (i.e., conclusions). For example, the text doesn't clearly state what these findings imply for future policy changes, how to improve the identified gaps, or why certain issues (like the overly bureaucratic process) persist, etc. In order for this and other text in the conclusions section to be considered conclusions, the text would need to go beyond describing the issues and start interpreting the significance of these challenges, such as explaining how they might undermine stunting reduction efforts and suggesting ways forward based on the findings (with reference to detailed recommendations as appropriate).
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 lessons – while brief and sometimes difficult to fully understand – appear to be rooted in the evaluation findings and likely have wider applicability and relevance beyond the object of the evaluation. However, without additional details, the reviewer cannot determine this.
ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	No	The ToR requested that the evaluation team identify lessons about what works and what does not work. The lessons learned and best practices section does not satisfactorily address this request. Additional details are required in order for the readers to understand each best practice and each lesson learnt. As an example, the lesson learnt from the impact of COVID-19 on stunting prevention programmes is unclear: 'There were opportunities to increase access to healthy foods for all families, increase awareness about the importance of healthy eating and to develop new programmes and policies that support healthy eating.' The operational bottlenecks highlighted in the efficiency findings could have been utilised to highlight/formulate lessons learned but were not. Lastly, the language of the first two paragraphs within the lessons learned section would have been more appropriate to include and expand upon as a conclusion.
SECTION G:	**RECOMMENDATIONS (weight 15%)**	63%	Comments on Rating
Question 17.	Are recommendations well grounded in the evaluation?		
i	Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.	Partially	Recommendations align with the evaluation purpose but are not always clearly formulated. The recommendations appear to be rooted in the findings of the evaluation. Detailed feedback on the formulation of the recommendations is provided in the feedback under question 17.ii, below, for better logical flow and completeness off feedback. Note that one major finding presented in the findings section is not reflected as a specific recommendation and it is unclear how or why this was missed: 'For the next strategic phase, starting in 2024, it will be important to have more SMART (specific, measurable, achievable, relevant and time-bound) indicators in a clearer results framework, as some of the indicators were vague and not clearly measurable.' A recommended action should have been logically derived from this finding.

ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	The recommendations provide some useful and appropriate recommended actions that have the opportunity to strengthen the national program. However, the recommendations section would have benefited from a revision to: 1) more clearly indicate the order of recommended actions as many appear to rely on another being completed first or alongside another both within and across pillars, 2) remove duplicative recommended actions as there are many actions across pillars that could be completed more efficiently in one exercise, 3) ensure that sufficient details are available to operationalise the recommended action, 4) ensure that the responsible party/ies is/are provided for each recommended action. Reorganising the recommendations to ensure that the order of recommended actions is clear and removing duplicative recommendations would have shortened the recommendations section and created a clearer guidance model for the intended users. As written, the recommendations section does not provide a clear roadmap for execution and the sheer number of recommended actions are likely to overwhelm the responsible party(ies) (there are over 60 recommendations). Most recommendations include the responsible party, with some exceptions. Some of these instances appear to be due to formatting (i.e., a recommendation is split into two bullet points). The recommendations section would have benefitted from a thorough review to ensure that designated party/ies were noted for each action and that recommended actions. Some recommendations do not contain sufficient details or recommended steps to operationalise. As written, it is difficult to discern how the recommendation should be executed. Some of the recommendations appear to feed into one another or require that they be executed in a specific sequence. However, the recommendations section is not organised in a way that allows for readers to easily understand the sequence in which recommendations need to be carried out. The recommendations could have been more clearly presented as step 1, 2, 3, etc. for clarity, noting the interdependencies of some 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 final report states that the recommendations draw on extensive, triangulated analysis of the key evaluation findings and conclusions, including the consultations with key stakeholders and a comprehensive document review. However, the process for developing the recommendations beyond using findings/conclusions from the evaluation is not provided. For example, the report does not note the involvement of duty-bearers or rights holders in the development of the recommendations. The report should have included a brief overview of the development process (e.g., whether the ERG reviewed and provided feedback).
Question 18.	Are recommendations clearly presented?		
i	Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.	Yes	Each recommendation states the stakeholders who would be responsible for carrying out the actions. The recommendations are prioritised according to whether they required immediate, medium, and long-term completion.
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	57%	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 report partially meets this criterion. The opening pages are missing the timeframe of the evaluation, date of report, commissioning organisation (Ministry of National Development Planning (Bappenas) and funded by UNICEF Indonesia) and location of evaluated object.
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 contain many of the items listed under this criterion, but the following are missing. --Terms of reference --List of site visits (note that this may have been due to confidentiality concerns). (Note also that Figure 14 in the body of the report gives an overview of which provinces and districts.) --Data collection instruments The report would have been strengthened by the inclusion of the additional annexes noted above.
Question 20.	Is the report logically structured?		
i	Structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles, well formatted).	Partially	The report uses numbered sections for each 'chapter'. Within each chapter, subtitles are formatted differently. However, it is sometimes difficult to understand the hierarchy of all of these subtitles and a numbered system would have allowed the reader to more easily follow along (e.g, section 1, 1.1, 1.1.1, etc.). There are also instances where sections are not properly formatted (e.g., sustainability in the table of contents falls under efficiency and gender, equity, and human rights section is not labelled in the table of contents, visuals are not properly numbered or referenced in the text, and the annexes are misnumbered (e.g., two annex 5s). The table of contents contains hyperlinks for easy navigation directly to a section from the table of contents.
ii	Structure accords to UNICEF guidelines for evaluation reports: context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.	Yes	The final report structure accords to UNICEF guidelines for evaluation reports.
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.	No	The evaluation ToR states that 'The draft report will fully conform to the Global Evaluation Report Oversight System (GEROS) of ideally 40 pages but not more than 70 pages plus executive summary and annexes that will be revised until approved.' The final report is over 100 pages exclusive of the executive summary and annexes.
ii	Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	The report is easy to understand and general free from grammar, spelling and punctuation errors. A final copyedit/proofread may have caught the remaining grammatical/punctuation errors, but these are few and far between and do not impact the reader's ability to consume the information.

iii	Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Partially	The report makes frequent use of visual aids which convey key information and add richness to the report. Nearly all of the visuals are clearly labeled. However, there are some exceptions to this. The following items lack a title but are still referred to in the text: --Figure 'defining a budget that supports the acceleration of stunding reduction' under the subsection 'budget' on page 9 --Figure of 11 key nutrition-specific interventions on page 17 --Figure of 9 nutrition-sensitive interventions on page 10 --Visual of members of the SUN academia and professional network on page 22 --Visual of first 1,000 days on page 41 --Visual on the five pillars on page 63 The following visuals/figures are titled but **are not** specifically referenced by figure/table number in the text: --Figure 15 on page 47 (Village Legal Authority, Regulations (Perbup) and Strategic Functions) --Figure 22 (Online Survey Results...Food Security?) on page 69 --Figure 27 (Government Funding...Stunting Reduction) on page 83 --Figure 28 (Human Development Cadres Roles and Responsibiliteis) on page 89 --Table 13 (National Programme...(KPIs)) on pages 110-112 There are also some instances where figures are referred to incorrectly (e.g., the text states 'in the figure below', but the figure is to the left of the text). This is likely due to final formatting. A final copyedit would have caught these inconsistencies. - Figure 10 lists the same title/name for both Phase 2 and Phase 3. - The report refers to Figure 15 in the paragraph/body of the text, but the figure it is referring to is titled Figure 14. - The report refers to Figure 18 in the paragraph/body of the text, but the figure it is referring to is titled Figure 15. - Table 2 is not referenced in the text but stands alone as the entire risks and limitations section. An introductory sentence or two prior to the table would have helped the document flow better, but it is not required. - Further, there is a list of figures and list of tables following the acronyms list for easy reference.
SECTION I:	**EVALUATION PRINCIPLES (weight 10%)**	71%	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.	Partially	The report states that the evaluation utilised a rights-based framework ('The methodological approach was also transparent, inclusive, gender sensitive and human rights-based, creating space for a diversity of voices and ensuring a particular focus across all aspects of the evaluation on the perspectives and experiences of those most at risk of exclusion.') The report does not specifically reference the CCC, CEDAW or other rights-related benchmarks in the design of the evaluation.
ii	Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).	Yes	The final report clearly describes the stakeholder types that participated in the evaluation as either key informants in data collection, members of the evaluation reference group (ERG) and overall management/quality assurance (UNICEF Indonesia and Bappenas). The evaluation management team is described on page 38 with members of the ERG noted in footnote 83. Stakeholders involved in data collection are detailed on pp. 27-28.
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 often uses language that is empowering and inclusive. However, the report also uses the term beneficiary/ies throughout. The authors should have used a more rights-based term other than beneficiary, such as 'participants' or 'rightsholders' to reflect the empowerment of individuals and communities in claiming their rights. The report clearly indicates the challenges encountered due to a lack of disaggregated data which prevents differential results from being assessed (see page 10: 'In terms of national programme coverage and disaggregation by gender, age and disability, the evaluation team was unable to find relevant information on the coverage and breakdown of these.') Where disaggregated data is available, the evaluation report notes this as well as its shortcomings (e.g., see findings under Pillar 4 and Gender, Equity and Human Rights). Due to Government of Indonesia standards, names of individuals involved in data collection are anonymised and show only organisation/departments (no information on gender, age or disability status is available).
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 investigated the level of inclusivity and equity of the programme (see gender, equity, and human rights sections of the report/evaluation matrix for further details). However, lack of data posed a major challenge. The evaluation report notes these challenges and puts forth recommendations to improve inclusive programming in future.
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)	6	
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.** Standalone section but also integrated into other pillars of the evaluation. c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? **YES.** A standalone criterion on gender, equity, and human rights was included in the evaluation framework. Issues of gender, equity, and human rights were also 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? **YES.**

ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily 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? **PARTIALLY.** --The evaluation describes the methodological approach in the inception report as 'gender sensitive'. However, details regarding how this approach was carried out during data collection and analysis are not available in the final report or its annexes. --Figures 18-23 discuss the online survey results, but these are not disaggregated by sex. Data from other data collection techniques (e.g., interviews, FGDs) are also not disaggregated by sex. --The evaluation methodology, methods and tools aimed to capture the differences in needs, experiences and impacts of the national strategy on people of different genders. The mixed-methods approach allows for a diverse range of stakeholder opinions and the purposive and random sampling also contributes to this diversity. The evaluation 'pillar' titled Gender, Equity and Human Rights explores the extent to which the programme is equitable and inclusive. However, the evaluation encountered many challenges regarding data availability. The evaluation offers recommendations for future actions to ensure that the strategy is implemented in an equitable and inclusive manner and that related data is captured. --The evaluation's data analysis technique of thematic analysis allows for gender analysis, and this is reflected in some evaluation findings (e.g., on social protection schemes under Coherence, data disaggregation of the BPNT programme under Pillar 4, antenatal care under Pillar 5, key challenges with nutrition-specific interventions, within the Gender, Equity, and Human Rights section, etc.). 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? **YES.** d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? **YES.** e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? **YES.** Formal IRB approval was also obtained.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	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? **NO.** 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.** However, this is largely due to the lack of available disaggregated data. c. Are unanticipated effects of the intervention on human rights and gender equality described? **NO.** 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? **YES.**

SWAP Rating Guidance

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