

UNFPA Evaluation Quality Assessment Grid

Version: March 2025 annotated

REPORT RATING SUMMARY			
Overall Rating		79%	Satisfactory
•••••	Excellent	5	
••••	Highly Satisfactory	4	
••••-	Satisfactory	3	The report meets UNFPA/UNEG standards for evaluation reports, but some indicators are inadequately addressed or missing. Decision makers may use the evaluation with some confidence.
•••--	Fair	2	
••---	Unsatisfactory	1	

REPORT DETAILS	
Title of the evaluation report	Country Programme Evaluation of the UNFPA Malaysia 1st Country Programme 2022-2025
Region	APRO
Country	Malaysia
Year of report	2025
Business Unit/programme country (managing evaluation)	Malaysia
Date of assessment review (dd/mm/yyyy)	10-Apr-25
Name of assessment review firm	IODPARC

CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered (list provided below)	1, 3, 5, 8, 18
UNFPA Strategic Plan areas covered (lists provided below)	
Three transformative results	Yes
Six outputs	Yes
Six accelerators	Yes
Organizational effectiveness and efficiency	Yes
Humanitarian evaluation	No
Evaluation evaluand (e.g. country programme/intervention/policy/thematic area)	Country Programme
Evaluation type (e.g. formative, summative, developmental)	Formative and summative
Geographic scope (e.g. global, regional, national)	National

EQA Summary: <i>The rater will provide top line issues for this evaluation relevant for feedback to senior management (strengths and weaknesses), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater</i>	
Strengths	
<ul style="list-style-type: none">Overall, this is a good-quality evaluation, broadly well-structured in line with the UNFPA Evaluation Handbook template.It is underpinned by a reasonably sound design and implementation.The analysis is also strong, drawing on a rich body of evidence that is generally well (though not comprehensively) referenced through footnotes or citations within the text.The report addresses all the analytical components outlined in the evaluation questions, making it comprehensive in scope and systematically aligned with the chosen analytical framework.The description of the UNFPA country programme is extensive, covering geographic, temporal, financial, and implementation status aspects and linking it to the the Malaysians socio-cultural context.Key stakeholders in the country programme are concisely and adequately described, with a useful table of implementing partners.The purpose, objectives and scope of the evaluation is clearly defined, outlining the reasons for the evaluation and identifying key intended users.The selected evaluation questions and criteria are appropriate for the purpose of the evaluation and are explicitly linked to the OECD/DAC criteria.A very extensive evaluation matrix is included, aligning evaluation questions with assumptions, indicators, and data sources, following the UNFPA Evaluation Handbook format.The report specifies adequate methods for data collection, analysis, and sampling, clearly describing both qualitative and quantitative data sources and a purposive sampling strategy.Evaluation findings are derived from a robust mix of evidence from a credible selection of sources, with evidence of triangulation.The conclusions and recommendation generally present a balanced perspective drawing from the evaluation findings to answer the evaluation questions and are linked to the preceding analyses and (in the case of the recommendations) have actionable guidance for implementation.	
Weaknesses	
<ul style="list-style-type: none">The most substantive weaknesses relate to some areas of superficial analysis—at times leaning more toward a descriptive list of UNFPA-supported activities, rather than offering a deeper examination of the outcomes for UNFPA's constituents in Malaysia.In several cases, the detail could have been streamlined, particularly where robust and well-referenced analysis already supports the key findings. Given the overall length of the report, a more selective approach, focusing on the most well-substantiated evidence, could have enhanced the report's quality and readability, even if this came at the cost of slightly softening some findings.The executive summary is missing an element and different components are unbalanced.The ToC is complex and difficult to follow due to its length and lack of narrative explanation linking outcomes, outputs, relationships, and causal factors.The ToC is insufficiently integrated into the overall evaluation logic; it lacks explicit connections to evaluation questions and assumptions.Rights-holders' voices are absent, limiting the evaluation's depth and inclusivity, despite logistical and contextual constraints.Ethical issues are only briefly addressed, with no dedicated section or annex providing substantive context or discussion.The use of ICT for remote data collection is not a meaningful innovation given its widespread use; evaluators should not present this as a novel approach.Some causal analyses are underdeveloped, especially on sensitive issues such as discrimination against LGBTQ+ individuals.Conclusion statements are inconsistently written, blending summative and recommendation-oriented language.Cross-cutting issues, especially intersectionality, are weakly integrated, with limited disaggregation of data by intersecting vulnerabilities due to CPD design and data limitations.	

Suggestions for future evaluators: <i>The rater will identify key suggestions to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, examples will be cited to assist evaluation managers</i>	
As noted above, the evaluation would have benefited from a more critical review to remove redundant, superfluous, or unsupported analysis, as well as overly descriptive elements that do not meaningfully contribute to, or in some cases, may even contradict, the evidence base. Such a review would not only have helped to reduce the report to a more manageable length, but also likely addressed some of the more noticeable editing errors, which detract from what is otherwise a potentially strong and credible piece of work. Some other areas for improvement are as follows:	
<ul style="list-style-type: none">The executive summary is missing an element and different components are unbalanced - the purpose/scope and methods sections, as well as the conclusions, could have been shortened to provide more space for findings.Some quite detailed descriptions or analyses (e.g. of the country programme and the CPD results framework) might have been more usefully included in an appendix.The theory of change, while comprehensive, is complex and somewhat challenging to grasp due to its length and lack of accompanying narrative description explicitly linking outcomes/outputs, relationships, and causal factors. Additionally, the TOC is somewhat "decoupled" from the overall evaluation logic as the explicit links with evaluation questions and/or assumptions were not undertaken.The absence of the direct voices of rights-holders is a definite limitation, although the evaluators deemed it not a likely avenue for research given the nature of UNFPA work in Malaysia and the inability to conduct field-based travel.The description of ethical issues and considerations is relatively cursory and does not substantively contextualise the principles or approaches. There is no specific section or annex on ethical considerations.The use of ICT for remote data collection is considered a marginal "innovation" given the widespread adoption of such tools – evaluators should be instructed that this does not qualify as an innovation.The underlying causal analysis for some questions/findings could be deeper and more nuanced, with limited analysis on sensitive topics like discrimination against LGBTQ+ people.The conclusion statements themselves have a number of writing/formulation issues, with a mix of summative and recommendation-type language.The integration of cross-cutting issues, particularly intersectionality, in the background and direct voice of different groups in the report, is limited. Data disaggregation by intersecting vulnerabilities is significantly constrained by the nature of the CPD and the limited quality of available secondary data.	

SECTION RATINGS			
SECTION A:	EXECUTIVE SUMMARY (weight 5%)	67%	Comments on Rating
Question 1:	Can the executive summary inform decision-making?		
	i Is a clear, standalone document useful for informing decision making, (a minimum of 5 pages, up to a maximum of 7 pages). <i>Note: YES - the executive summary is within the indicated maximum page limit. PARTIAL - the executive summary exceeds the maximum page limit by 1 to 2 pages. NO - the executive summary exceeds the maximum page limit by more than 2 pages</i>	Yes	The executive summary spans six pages, which is appropriate given the size of the report. It is clearly distinguished from the main report.
	ii Includes all necessary components of the evaluation report, including: (1) overview of the context and intervention, (2) evaluation purpose, objectives and intended users, 3) scope and evaluation methodology, (4) summary of most significant findings, (5) main conclusions and (6) key recommendations	Partially	Five of the six mandated components are present in the executive summary. The summary does not have an overview of the context and CPD - it goes straight to the purpose/scope of the evaluation.
	iii Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Partially	As noted above, the summary misses the first component - a description of the overall context of UNFPA in Malaysia within which the evaluation is situated. In addition, the findings section in the summary is very short - less than one page. The evaluation conclusions and recommendations are reproduced almost in full, and thus comprise most of the evaluation. The purpose/scope and methods sections could have been shortened, as could the conclusions, to provide additional room for the findings.
SECTION B:	BACKGROUND (weight 5%)	80%	Comments on Rating

Question 2.	Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?		
	i Clear description of the evaluand (e.g. intervention), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes	<p>There is an extensive description of the UNFPA country programme provided in section 3.2. This covers geographic, temporal (implementation period), financial and implementation status aspects of the country programme. Importantly, the report notes that programmatic activities largely commenced in the latter part of 2023 when funds became available, and many were still in progress at the time of the evaluation cut-off.</p> <p>While comprehensive (particularly the financial description), such a level of detail might have more usefully be included in an appendix, with a more concise summary of the salient points presented in the main report. Although the length does not affect the rating of this criterion, report length considerations are a concern for others, so this is one section that could have been truncated without impinging on report quality.</p>
	ii Clear description of the context of the evaluand (e.g. economic, social and political context, relevant aspects of UNFPA's institutional, normative and strategic framework, cross cutting issues such as gender equality and human rights, disability and LNOB dimensions) and how the context relates to the evaluand (e.g. key drivers and challenges that affect the implementation of the intervention/policy/thematic area	Partially	<p>The context of the UNFPA country programme in Malaysia in section 2 explicitly addresses the economic, social/demographic and developmental contexts. While the political context is not explicitly detailed as a separate subsection, references to government strategies such as the twelfth and (upcoming thirteenth) Malaysia Plans offer insights into the national political landscape and priorities. The report also links the context to the UNFPA country programme, noting the sensitive nature of UNFPA's work in areas like SRHR and gender within the Malaysian socio-cultural and religious context. This is all highly relevant and useful detail.</p> <p>This said, some of the population statistics cited are somewhat dated, for example contraceptive usage data is from the 5th Population & Family Survey (MPFS-5) which was in 2014. While MPFS-6 was still underway at the time of research, some more recent source (even if less authoritative) would have been more useful. Furthermore, there is no significant description of the LNOB and disability inclusion context in Malaysia, beyond reference to these within the UNFPA programme plan. The UNFPA SP 2022-2025 focuses on finishing the unfinished business of the ICPD and (importantly) reaching the furthest behind first. This is also an essential accelerator of the UNFPA SP 2022-2025, hence the 'partial rating' as it is not well covered in an otherwise well-written section.</p> <p>Finally, the background to UNFPA work in Malaysia includes some useful background to the country programme under evaluation, providing some information on the prior activities of UNFPA as well as describing the main findings of the evaluation of the programme plan - useful context as they guided the priorities of the evaluand.</p>
	iii Linkages drawn between the evaluand and the ICPD benchmarks and SDGs relevant targets and indicators.	Yes	<p>The report includes a brief narrative summary of Malaysia's progress with respect to the SDGs, presented in section 2.1 (para 13), where SDGs 3 and 5 are briefly discussed in terms of their performance against targets (the progress is summarised rather than itemised, all narratively). There is another reference in section 2.2 (para 17), again briefly and not itemising any specific achievements or targets.</p> <p>The report also references the SDGs in other sections, such as noting alignment of the country programme with international frameworks like the ICPD Programme of Action and the SDGs, and providing a table (Table 3.2) mapping CP outputs to ICPD-PoA priority areas (as well as UNFPA global outcomes, UNSDCF outcomes and national plans) and noting the relevant SDGs - albeit very briefly (i.e. listing the relevant SDGs in the table only). The report simply presents these linkages without detailed critical analysis of specific indicators and/or targets. The report does include a 'key facts' table in the opening pages (page xi) as per the guidance in the UNFPA Evaluation Handbook, but this does not include any explicit reference to the SDGs, although does have information on specific population indicators that are integral to the SDGs.</p> <p>The description of the evaluand (i.e. UNFPA's programming in Malaysia) also notes linkages to relevant SDGs.</p>
Question 3.	Are key stakeholders clearly identified and analysed?		
	i Clear identification of key stakeholders which should include implementing partner(s), development partners, rights holders, and duty bearers among others; and of linkages between them (e.g., stakeholder map).	Yes	<p>As discussed in 2i above, the main stakeholders in the country programme are concisely and adequately described in section 3.2. The report explicitly lists partners implementing interventions since the start of CP1 in 2022, usefully (and concisely) in a table (table 3.6, page 29). The different government ministries responsible for coordinating activities related to UNFPA's mandate are described in section 2, with useful additional information provided in table 3.7 related to the funding mobilised by UNFPA directly from the Government of Malaysia - a key area of interest for this evaluation as this is atypical for UNFPA operations and hence warrants particular mention.</p>
	ii Stakeholders are analysed to understand their specific rights, duties, needs, interests, concerns, and potential impact on the evaluand.	Partially	<p>The report does not provide extensive analysis of the various interests, needs etc. The analysis/mapping is limited to data related to UNFPA funding and some specific areas of intervention in the narrative.</p>
SECTION C:	EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	100%	Comments on Rating
Question 4.	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	<p>The four elements of the purpose of the evaluation are clearly defined in Chapter 1 of the report. It clearly notes that the CPE is required according to UNFPA's evaluation policy and the Country Programme Document, which mandates an evaluation in the penultimate year of the programme cycle (2022-2025), hence the evaluation covering the period January 2022 to June 2024 is exactly as per requirements. The report explicitly outlines the four main purposes per UNFPA evaluation guidelines. The key intended users are also clearly identified: decision-makers within UNFPA Malaysia and other offices, government counterparts, national stakeholders, the UNFPA Executive Board, other development partners, UNFPA APRO and Headquarters, and the UN Country Team in Malaysia. This is clear and unambiguous.</p> <p>A minor suggested improvement would be to explain why this is the first country programme for UNFPA Malaysia. As the preamble to Chapter 1 states that UNFPA has been active in the country since 1973, the reason for the absence of previous CPs could be noted, maybe via a brief footnote.</p>
Question 5.	Are the objectives and scope of the evaluation clear and realistic?		
	i Clear and complete description of the objectives of the evaluation, including reference to any changes made to the objectives included in the ToR (if applicable).	Yes	<p>Although not presented in a separately labelled section/subsection, the report nonetheless clearly outlines the three specific objectives of the evaluation on paragraph 5. There are no changes to the <i>specific</i> objectives from the ToR, although there are changes to the evaluation questions (fully explained in the report in section 1.3 - see 6.1 below).</p> <p>However, there is some deviation from how the TOR articulates the evaluation objectives - essentially the TOR presents two sets of objectives: two high-level objectives and three specific objectives. This is somewhat confusing in the TOR, and the evaluators have ameliorated this by including the text of the two high-level objectives in the preamble text (para 4) but not explicitly labelling them as such, and focusing attention on the three specific objectives. This is an acceptable and useful clarification.</p>
	ii Clear and relevant description of the scope (e.g. thematic, geographic, and temporal) of the evaluation, covering what will and will not be covered, 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, humanitarian or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	<p>The scope of the evaluation is presented in section 1.2. The description explicitly notes the:</p> <ul style="list-style-type: none"> - thematic scope (i.e. the UNFPA Strategic Plan's output areas and cross-cutting issues and programmatic areas of SRH, adolescents and youth, PD and GEWE/GBV, and also including some operational aspects of the CP); - geographic scope (focused on Kuala Lumpur but also other geographic locations based on the reach of UNFPA Malaysia's interventions); and - temporal scope (January 2022 to June 2024). <p>While all three dimensions of the scope are thus clearly and concisely described, there is a slight inconsistency in the temporal scope in that the limitations section (1.3.4) notes that UNFPA Malaysia operations did not begin until mid-2023, so hence the <i>de-facto</i> temporal scope was from this time. This might have been mentioned in the section for clarity.</p>
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	86%	Comments on Rating
Question 6.	Are the selected evaluation questions and evaluation criteria appropriate for the purpose of the evaluation and is there clear justification for their use?		

	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 evaluation questions and sub-questions, presented in section 1.3 appear appropriate for meeting the objectives and purpose of the evaluation. The evaluation is explicitly designed around the OECD/DAC evaluation criteria of relevance, coherence, effectiveness, efficiency, and sustainability (noted in the purpose and objectives). Table 1.1 explicitly lists the EQs in line with the specified OECD/DAC criteria. The report states that the evaluators reviewed and refined the EQs proposed in the ToR, ensuring they covered these five criteria, and agreement on the final eight EQs was reached with the CO and endorsed by the ERG. While the evaluators note that the "core premise" of the original EQs was maintained, the merging of some EQs constitutes a change from the initial ToR, and this is explicitly mentioned in the report (with individual changes explained via a footnote) - this is fully in line with good evaluation practice.
	ii	Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding evaluation questions, indicators, lines of inquiry, benchmarks, assumptions, source of data, methods for data collection and analysis, and/or other processes from which the analysis can be based, and conclusions drawn.	Yes	There is a very extensive (46 pages) evaluation matrix is presented in Annex 1. It presents the EQs as per the formulation in Table 1.1, along with varying numbers (1-4) of key assumptions to be tested, indicators related to each, and a list of data sources (specific individuals or organisations, organized by whether the data source is primary/secondary and the collection tool) with specific examples of data collected (linked to the assumption and corresponding to the indicators). This is exactly in accordance with the format and template mandated by the UNFPA Evaluation Handbook. The assumptions and indicators are well derived from the evaluation questions and, if fully addressed, should provide a comprehensive database of evidence that can robustly address each evaluation question. The selection of data/evidence presented against each of the assumptions and data collection tool is very comprehensive, relevant and robust.
Question 7. Is the theory of change, results chain, logical framework, or equivalent framework 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 evaluation is explicitly specified as being theory-based in the description of the evaluation approach (Section 1.3). This section describes how the different elements of the intervention logic of the country programme, including the theory of change, was utilised by the evaluators as the primary point of reference for development of the findings, conclusions and recommendations. The theory of change itself is then presented in section 1.3.1 (in table/graphical format) following the programme priority areas as articulated by the CPD. The description of the intended results of the CP and how the evaluators have presented them in the TOC, is clear and comprehensive.
	ii	Causal relationships between the various elements (e.g. outcomes, including the three or relevant Transformative Results, outputs) of the theory of change, results chain or logical framework are presented in narrative and/or graphic form).	Partially	As noted, the intervention logic/theory of change is presented in graphical/tabular format in table 1.1, section 1.3.1. This is quite comprehensive, although is complex and somewhat challenging for the reader to grasp easily as: a. The table is long - it covers almost three pages which makes it difficult to read. b. It is not accompanied by a narrative description of the outcomes/outputs, relationships and causal factors. The evaluators could have improved this section by devising a more simplified graphical representation (on one page) and accompanying this with narrative which covered much of the detail that is currently present in the ToC. For example, many of the sub-outputs in the table could have been removed to narrative. The ToC also misses key elements such as a problem statement and assumptions about causal relationships (why change happens), contextual factors (under what circumstances change happens), and stakeholder characteristics (who makes change happen (or not) and for whom). A complete ToC should include these elements.
	iii	Comprehensive analysis and assessment of the theory of change, results chain or logical framework, and if requested in the ToR, it is retrofitted/reconstructed by the evaluators.	Yes	The theory of change as presented in the evaluation report was developed by the evaluators based on the results framework and accompanying narrative of the CPD under evaluation. This was specified in the evaluation terms of reference as a critical step in the evaluation, as a review of the validity and comprehensiveness of the theory of change was deemed by UNFPA Malaysia to be crucial to inform the next country programme's theory of change. The evaluators do present some initial judgement on the validity of the TOC in section 1.3.1 (paras 14 through 17), which is certainly useful, albeit brief. The evaluators do, however, refer to a more granular analysis to take place as part of the findings section - specifically under EQ3, which explicitly covers the TOC. While the criterion for the rating has largely been met (and is rated as such), the additional good practice of explicitly linking the elements of the TOC with the evaluation questions and/or assumptions has not been undertaken, thus the TOC is somewhat 'decoupled' from the overall evaluation logic itself. As the TOC is specifically analysed under EQ3, this is less of an issue than if such analysis was not mandated.
Question 8. Does the report specify adequate methods for data collection, analysis, and sampling?				
	i	Evaluation design and set of methods are clearly described, and are relevant and robust for the evaluation's purpose, objectives and scope, including the use of AI in the evaluation process if applicable.	Yes	The approach of the evaluation is clearly explained as being theory-based, and this is quite evidently appropriate to the nature of the work, and fully in line with the approach specified in the TORs. Section 1.3.2 provides a clear and relatively concise description of the evaluation design and set of methods, with the purpose, objectives, scope (thematic, temporal, geographic), and evaluation approach explicitly laid out across the entire of section 1. The eight evaluation questions are clearly aligned with the evaluation criteria being applied (relevance, coherence, effectiveness, efficiency, sustainability) and the overall objectives. The description of the data collection methods is mostly clear (with one caveat - see below) and are suited to the context and should allow a diversity of perspectives to be gathered, facilitating triangulation between findings (albeit without the direct voices of rights-holders). The report explicitly states that AI tools or technologies were not used in the evaluation process.
	ii	Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR).	Yes	The evaluation report clearly describes both qualitative and quantitative data sources in sections 1.3.2 and 1.3.3. One minor issue is that the description of the methods on page 8 of the report (paras 19 and 20) places the data sources under the headings of Quantitative and Qualitative data - they would be better described as "secondary" and "primary" sources, as documents/datasets etc. cited in para 19 include both quantitative and qualitative data. This said, the (secondary) quantitative data sources primarily consisted of UNFPA reported programmatic and financial data, national datasets/population surveys, census reports, and implementing partner reports with an aim to collate data on key indicators. The (primary) qualitative data specified was either observation or semi-structured interviews with both internal and external stakeholders.
	iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Yes	The evaluation report specifies a purposive sampling strategy for data collection from both secondary and primary sources (described in section 1.3.3). This involved selection of UNFPA interventions (with a full list provided in Annex 5) via a filtering exercise and iterative stakeholder mapping with the UNFPA Malaysia CO to identify the major categories of stakeholders distributed across the CP themes and covering all outputs. The evaluators applied a number of useful criteria in selecting the specific participants to be approached/included, directed by UNFPA evaluation guidance to minimise bias. The stakeholder sampling was focused at the national level, as most programme interventions were focused on national-level (advocacy, policy dialogue, partnership, coordination, and capacity building). However, it also included selected stakeholders at the sub-national level in Sabah, Sarawak, and Selangor who participated in or were involved in interventions. The report has a useful table of the sampling framework (Table 1.2) which details the initial sample targets and the final result, with 52 of 68 individual approach ultimately participating. Annex-4 provides a listing of the final 52 stakeholders interviewed. In terms of representativeness and diversity of perspectives, the described data sources appear overall relevant and robust for the evaluation, but the absence of the voices of rights-holders is a definite limitation. The evaluators did seek some direct inclusion of these, but the absence of field-based travel and the nature of UNFPA work in Malaysia (not community-oriented) described in the limitations section meant that the evaluators deemed it not a likely avenue for evaluation research.
	iv	Methods allow for rigorous testing of the theory of change, results chain or logical framework (e.g. methods help to understand the causal connections, if any, between outputs and expected outcomes (3TRs).	Yes	While the evaluators state that the evaluation methods described in the report were explicitly designed to test the theory of change and results chain through the theory-based approach and contribution analysis, as noted above, the link between the elements of the TOC and the evaluation matrix/evaluation questions could have been more clearly noted. This said, the evaluation matrix clearly and comprehensively links data collection to evaluation questions via the assumptions and indicators. Effective application of the elements of the matrix should facilitate good data triangulation between quantitative/qualitative/primary/secondary data. The reliance of the evaluation on secondary programme data to validate the TOC may be a challenge given that: a. The timeframe covered by the evaluation is limited; b. There was limited regular monitoring explicitly linked to the TOC during programme implementation, and c. UNFPA/implementing partner monitoring data tended to focus more on outputs/activities. Despite these limitations (which are mostly described - with mitigation strategies - in 1.3.4), the chosen methods provided a considerable basis for understanding the causal links between UNFPA's outputs and expected outcomes.

	v	Clear and complete description of the methods of analysis, including explainability and full disclosure of the use of AI in the evaluation process, if applicable.	Yes	The evaluation report provides a description of its methods of analysis in section 1.3.2. It clearly states that the evaluation matrix was used as the basis for data collection and analysis. This involved synthesis and triangulation of information obtained from the document reviews and key informant interviews. The section notes the different approaches to analysis employed by the evaluators, including descriptive, thematic, content, comparative, quantitative, and contribution analysis. The section also usefully describes the process of coding of interview transcripts and documents, collaborative analysis among team members, and triangulation across sources and methods to maximize validity and reliability. As noted above, the report clearly notes that AI was not used by the evaluators.
	vi	Clear and complete description of limitations and constraints faced by the evaluation in its data collection and analysis, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Yes	Section 1.3.4 of the evaluation report is a dedicated section on limitations and mitigation measures. It explicitly identifies various constraints faced during the evaluation process, including on timeframes, field travel to engage with stakeholders directly, the absences of rights-holders and limitations on secondary quantitative data. For each of these identified limitations, the report describes mitigation measures undertaken, such as use of a mix of different primary and secondary datasets and conduct of online interviews with stakeholders in Sabah and Sarawak. While the report notes that data limitations were addressed through efforts to triangulate across sources and methods, the report acknowledges some gaps in evidence, such as the inability to obtain the Ministry of Health's perspective through primary data collection, although this was partially filled by their participation in ERG meetings and written feedback. This, however, was not explained in the limitations section, but in the sampling section (1.3.3) - the text could have been moved. The description of the limitations, corresponding mitigation measures, and (importantly) acknowledgement of remaining evidence gaps warrants rating this criterion as "yes", but it could have been enhanced by including the limitation noted above regarding sampling, and also presenting the section in a more structured format - the narrative/paragraph style is somewhat difficult to parse. Use of a table with limitations, mitigations and outcomes in separate columns is a useful presentation format that is more easily understood.
Question 9.		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 UNEG Ethical Principles.	Partially	While the evaluation report makes explicit reference to the UNEG Norms and Standards and UNEG Ethical Guidelines for Evaluation in section 1.3 (para 11), the description is relatively cursory and does not substantively contextualise the principles or approaches. There are clear statements that the CPE adhered to UNEG Norms and Standards and followed the methodology prescribed in the UNFPA Evaluation Handbook, and notes that the evaluation was "ethical, transparent, inclusive, and participatory", and followed the "do-no-harm" approach". The data collection section (1.3.2) provides some additional detail on ethical approaches, including the use of informed consent procedures "as per UNEG guidelines". There are some further details, although they are brief, e.g. "laying out ground rules for interviews" (although these rules are not described - they may refer to the "sequencing of interviews" steps itemised in Annex 6 which has the data collection tools). While some other measures to ensure confidentiality and anonymity of data (via encrypted storage and anonymization techniques) are described, overall there is not a great deal of detail. While the absence of rights-holders in the evaluation means that the ethical bar is not as high as it would otherwise be, some more contextualisation of the UNEG obligations and/or principles is warranted.
	ii	Clear description of ethical issues and considerations (e.g. respect for dignity and diversity, fair representation, confidentiality, and avoidance of harm) that may arise in the evaluation, safeguard mechanisms for respondents (e.g. parental consent forms for adolescents, compliance with codes for vulnerable groups; WHO standards of safe data collection on GBV) and ethical considerations in the use of AI as applicable (e.g., transparency of use, explainability, privacy, data protection, accuracy, human rights). If AI is used in the evaluation, there should be transparency and disclosure on the ethical and	Partially	As noted above, while the report has some description of ethical issues and considerations, i.e. explicit statement of adherence to UNEG Norms and Standards for Evaluation in the UN System and UNEG Ethical Guidelines for Evaluation, confidentiality assurance, verbal informed consent before each interview, ensuring voluntary participation, and allowing participants to decline to answer or stop the interview at any time, the references are brief and not well-contextualised for the specifics of the overall evaluation. The data collection tools are annexed, and include measures to ensure ethical data collection, and the data management processes are reasonably well described. However, there is no specific section on ethical considerations nor is there an annex on this. As noted, while the evaluation focused exclusively on institutional stakeholders, and thus more stringent ethical approaches were not warranted, the level of detail as it stands warrants only a "partially" rating.
Question 10.		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 include efforts to optimize the evaluation process (e.g., use of AI or new technology for data gathering, content analysis, outcome harvesting among others), or components introduced to enhance inclusion and participation in the evaluation processes (e.g. a youth steering committee), or ways of sharing of evaluation results.	Partially	The TOR notes that "innovative and context-adapted evaluation tools (including ICT)" are encouraged for the evaluation. The evaluators did make use of ICT for remote data collection among stakeholders in eastern Malaysia (Sabah, Sarawak and Selangor), which warrants the "partially" rating, but this is on the margin on innovative approaches, given the widespread adoption of such tools in the past decade for evaluation and similar research. No other description of innovations in terms of technology or evaluation practice are described.
SECTION E:		EVALUATION FINDINGS (weight 25%)	83%	Comments on Rating
Question 11.		Do the findings clearly and adequately address all evaluation questions and sub-questions?		
	i	Findings are presented clearly and provide sufficient levels of evidence to systematically address all the evaluation's questions	Partially	The presentation of the findings is largely, if not fully, in accordance with UNFPA evaluation guidance, i.e. they are structured around the eight evaluation questions, they include a brief summary of the overall tenor of the findings against the evaluation question. This summary is not quite in line with the evaluation guidance, insofar as the summary (in some cases, e.g. very much EQ1, less so EQ2) is essentially a restatement of the findings under the question, rather than a more nuanced summary of the analysis and findings combined - and there is no reference to corresponding evidence in the evaluation matrix. All findings are numbered sequentially. One evaluation question (EQ2) does not correspond fully to the original version - the text in the findings omits the reference to "national priorities, strategies and policies". The actual analysis and findings underneath this, however, do include clear consideration of this element, so it is likely an omission in the EQ text only, rather than a missed point of analysis by the evaluators. Finally, the structure of the analysis is not fully in accordance with the guidance here - many of the findings are couched in terms of 'conclusion'-type statements, even with language that is more related to recommendations (e.g. Finding #2: "Thus, further efforts to reach those not yet reached will be needed through..."; Finding #5: "UNFPA programming needs to keep in view the emerging trends..."; Finding #10: "To overcome gaps identified, the TOC for the next CP would benefit from being strategically developed more holistically for the full programme with a more integrated approach", Finding #40, etc.). The language is not in all places fully an objective presentation of evidence that speaks to the finding but sometimes unsubstantiated judgement-type language that talks about what UNFPA "should" do (i.e. a recommendation).
	ii	Explicit use of the evaluator's theory of change, results chain, logical framework in the formulation of the findings.	Yes	There is explicit use of both the reconstructed/amended TOC and the UNFPA Malaysia results framework in the development of the findings. Assessment of the validity of the TOC is specifically covered under EQ3 (relevance), while a detailed reconciliation of the results achieved against the original CPD outcomes and outputs is covered under EQ5 (effectiveness). There is a specific table of these results presented in this section (pages 47-50) which, while important analysis to have covered, would have been better placed as an annex, with a summary of the level of achievement (or otherwise) of the various indicators/targets presented in the main report with discussion of the implications of this for UNFPA effectiveness.
Question 12.		Are evaluation findings derived from credible data sources as well as a rigorous data analysis?		
	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 using multiple data sources.	Yes	The evaluators have used a robust mix of evidence from a credible selection of sources for generating findings. This is largely qualitative data, primarily through extensive key informant interviews with a diverse range of stakeholders from government, implementing partners, UN agencies, academia, and civil society. The evaluation also incorporates some quantitative data through the presentation of progress against defined CPD output indicators, as shown in Table 4.1, as well as references to financial data and some secondary data sources. The findings clearly present data relevant to both output and outcome levels of the UNFPA country programme. Progress towards specific CPD output indicators is tracked quantitatively, while the discussion of findings relates these outputs to the broader intended country programme outcomes, such as improved SRHR access and gender equality. Further, triangulation is evident in the report through the comparison of information obtained from different data sources, including stakeholder interviews and programme documents/datasets. For example, under EQ4 (coherence), perspectives from multiple types of external stakeholder (via KIs) - government, sister UN agency, implementing partner - are coupled with analysis of financial data relating to government contributions to UNFPA's budget to build a well-triangulated case for the finding.

	ii Findings are clearly supported by the evidence presented, 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 are generally supported by the evidence presented, although the strength of the support varies across different evaluation questions. The findings rely heavily on substantiation on evidence from KIs with various stakeholders, which are sometimes directly quoted, which is useful and adds readability to the report. The findings do draw upon desk/document review data, for example Annual Work Plans, the CPD (and associated results framework) as well as financial data and other relevant reports, both internal and external. For instance, Finding #1 on the CP's adaptation to diverse populations mentions engagement with stakeholders and provides an example of a needs analysis conducted by NPFD8. Similarly, Finding #3 on alignment with frameworks references the 12th Malaysia Plan and SPV 2030. A range of external studies, reports and publications (some of which are 'gold-standard' such as the findings of the UPR) are also cited in findings, adding to the credibility of findings. The extent to which the findings are based on clear performance indicators, standards, benchmarks, or other means of comparison also varies, depending on the evaluation question - which is to be expected. Table 4.1 provides a clear overview of the progress towards CPD output indicators, including baselines, targets and progress, offering a direct quantitative comparison. However, for many qualitative findings, the basis for assessment relies more on the perspectives and opinions of stakeholders gathered through interviews and the evaluation team's analysis of programme design and implementation. For instance, the assessment of UNFPA's strategic partnerships (EQ4) is largely based on the endorsement of UNFPA as a strategic partner by key informants. There are some unsubstantiated sub-findings (e.g. Finding #21, para 77), and the report would have benefited from a more critical review to shorten some of the more descriptive text that simply lists activities, but overall the recourse to evidence is good. The evaluation does present both positive and negative findings, highlighting both successes, such as the development of anti-sexual harassment laws, and areas needing improvement, like the application of the "leaving no one behind" principle. The report acknowledges slow progress on some indicators due to delayed funding and identifies challenges like the lack of gender and social inclusion M&E frameworks (under Finding#29).
	iii Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or not- from outputs to high level results).	Partially	There is partial identification of underlying causal factors across the report. For example, EQ5 (effectiveness) does explicitly itemise the enabling and inhibiting factors in the implementation of interventions, and the findings under this question discuss factors such as the late signing of the CPAP and the resulting late availability of government funding as a major reason for slow progress across several outputs. Human resource constraints and frequent changes in leadership are also mentioned as impacting the continuity of stakeholder engagement and timely delivery. The evaluation also highlights the lack of a multi-sectoral approach and limited teacher capacity as inhibiting factors for mainstreaming CSE. Conversely, strong partnerships with government agencies are identified as an enabling factor for achieving goals. The logical chain from outputs to higher-level results is covered largely with respect to the existing CPD results framework (i.e. under EQ5, effectiveness). The EQ on the TOC (EQ3) is focused mostly on the overall relevance/validity of the TOC at a high level and does not explore the logical progression (and, at two pages, is quite brief). For some questions/findings, the underlying causal analysis could be deeper and more nuanced. For example, under Finding #13, the added-value of UNFPA partnership with the Government around SRHR issues is highlighted due to the "sensitive operating environment when it comes to SRHR and gender issues" (para 48, page 43) - this could have been unpacked further as the ability of UNFPA to balance these "sensitivities" with its mandate to ensure SRHR for all, including vulnerable groups, is vital. There is some more useful analysis under EQ5 (effectiveness), which explores the dynamics of SRHR for youth, although the interesting analysis is somewhat lost in a considerable amount of descriptive text around activities. In particular, there is limited analysis of UNFPA efforts to address discrimination against LGBTQ+ people (who are criminalised in Malaysia) - while there are a handful of references, the evaluators appear to steer clear of unpacking these potentially controversial topics but that are still key elements of UNFPA's mandate.
Question 13. Does the evaluation assess and use the intervention's Results Based Management elements?			
	i Assessment of the adequacy of the intervention's planning, monitoring, and reporting 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 TOR for the evaluation specifies the monitoring and evaluation of the country programme as one of the thematic areas of scope, hence its assessment is warranted. The report does explore RBM elements to an extent - specifically, finding #40 under EQ7 (efficiency) addresses the systems used by UNFPA (the SIS and Quantum) for activity-level monitoring, programme monitoring. It delivers some critical analysis in this regard, notably the orientation of these systems towards "activities performed at the IP level, rather than the results and outcome levels" of those activities, as well as other issues related to the system. This is all useful, although some unpacking of the specific criticisms would have been good. For example, the evaluators note that it is not clear how indicators at IP level contribute to CP output indicators, indicating a potential gap in the vertical logic of the results framework, as well as different interpretations of the information to be included in progress reports among IPs. This is all interesting and potentially useful, but some more specific examples and concrete evidence would be of more value to UNFPA. For example, finding #40 notes that "there was seen to be a lack of reporting on the progress and achievement of project indicators" - but no more specific analysis of this is provided, such as a quantification of the missing data, or even specific examples to illustrate the point. Overall, while the criterion has been met, it would have benefited from some more analysis.
SECTION F: EVALUATION CONCLUSIONS (weight 10%)		50%	Comments on Rating
Question 14. Do the conclusions clearly present an unbiased overall assessment of the evaluand?			
	i Conclusions are clearly formulated and present unbiased summative statements that respond to the evaluation questions.	Partially	The ten conclusions are generally clearly articulated, following a structured format with evidence links, and, notwithstanding an issue with how the conclusion statements themselves are formulated (discussed under the next criterion), present a balanced perspective drawing from the evaluation findings to answer the evaluation questions. They are linked explicitly to the evaluation criteria, EQ and related recommendations (although not the specific findings, which, given that there are 47 findings in the report, would have been useful) and are organised according to strategic and programmatic levels. The statements themselves are mostly clear, although some of the conclusions headline statements have a significant issue with their formulation, which is more related to recommendations than to conclusions. For example, conclusion 1 has a headline sentence which is good, but is followed by what appears to be a recommendation, although it is ambiguous due to a writing error with the text reading "To drive further progress, the country's capacities [should?] be leveraged for broader regional impact." The rest of the conclusion speaks about what should happen which is a recommendation. Conclusions 1 through 5 have this mix of summative and recommendation-type language. Conclusion 6 is more appropriately worded, but 7 through 10 again speak about what is needed rather than what has been found. While a significant issue, it is largely semantic, in that the statements could be reformulated to speak to what has been found by the evaluators, rather than the needs that these findings lead to.
	ii Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Partially	The conclusions generally go beyond simply rephrasing individual findings, i.e. they synthesise related findings to provide a more in-depth analysis of the UNFPA country programme. For instance, Conclusion 1 draws upon the discussion of the CO adaptation to the UMIC context (Evaluation Question 6) to offer a broader insight about strengthening financial models, intellectual leadership, policy influence, and partnerships, and leveraging country capacities for regional impact. Similarly, Conclusion 3 regarding the use of resources originates from multiple evaluation questions (4, 5, and 7) and synthesises findings related to coherence, effectiveness, and efficiency to provide an overall assessment of resource management despite limited capacity. However, while the conclusion synthesizes the findings, they do not necessarily elevate the analysis and provide a judgement of the country programme itself. In that sense, the conclusions are somewhat overly-summative and duplicative of the findings. In addition, there are several conclusions (conclusions 2, 6, 7, 8 and 9) that draw on a single evaluation question and its findings only. Quality conclusions should cut across several findings across diverse evaluation questions, identifying issues that emerged in the responses to many evaluation questions.
Question 15. Are lessons learned identified? [N/A if lessons are not referenced or requested in ToR]			
	i Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	Although part of the purpose of the evaluation as stated in the TOR is to "generate evidence and lessons", the TOR does not specifically note a need for "lessons-learned" in the structure or framework of the report, hence this criterion is not rated. Nonetheless, the evaluators do include a short lessons-learned/good practices subsection (pages 78/79). This presents five interesting, useful and actionable/replicable practices and lessons that are worthy of being highlighted - these have wider applicability also, for example the importance of engagement with religious leaders, which is a practice that has been used by UNFPA to good effect in other country contexts also, so worthy of note here. There is some overlap with the "inhibiting and enabling factors" subsection under the Effectiveness criterion (pages 60-62), although not fully. Should the evaluators have sought to reduce the overall length of the report, they could have combined both subsections and eliminated duplication.
	ii Lessons learned are clearly presented and provide actionable insights on the positive aspects of the evaluand as well as any areas of improvement.	Not Rated	Although part of the purpose of the evaluation as stated in the TOR is to "generate evidence and lessons", the TOR does not specifically note a need for "lessons-learned" in the structure or framework of the report, hence this criterion is not rated.
SECTION G: EVALUATION RECOMMENDATIONS (weight 15%)		88%	Comments on Rating
Question 16. Are recommendations well-grounded and articulated?			

	i Recommendations are clearly formulated and logically derived from the findings and/or conclusions.	Yes	The formulation and presentation of the seven recommendations is clear - they are quite clearly written, with a headline statement following by sub-recommendations. There is clear and explicit linkage between the recommendations and the relevant conclusions (the specifically-referenced conclusions are named by number) and the logical derivation is clear.
	ii Recommendations are useful and actionable for primary intended users. Specific guidance is provided for its implementation (e.g. actions, deadlines, responsible actors), as appropriate.	Partially	The recommendations presented in Chapter 6 (pages 85-90) are useful and actionable for the primary intended users, with specific guidance for the implementation of each provided through a series of "operational implications". These are analogous to sub-recommendations or action points and offer practical considerations for implementing the recommendation. For example, Recommendation 1 suggests specific types of partnerships to pursue (Parliamentarians, private sector, Human Rights Commission). Recommendation 2 outlines practical and detailed measures for strengthening the M&E system. This is quite positive and robust. One issue with the recommendations is the division between "strategic", "programmatic" and "operational" recommendations (three strategic, two programmatic, one operational). While such a categorisation is useful and common practice in evaluations, the introduction to the recommendations notes that some are "quick-fixes" for the current CP, while others are "more forward-looking for incorporation in strategies to be devised for the new CP". This is contrary to the strategic vs. programmatic/operational categorisation - strategy, by definition, is generally more long-term than immediate operational adjustments. Further, the operational implications under all are a mix of operational/programmatic and strategic elements. Indeed, the final recommendation, deemed 'operational' by the evaluators, is around investment in human resources for building UNFPA's profile in Malaysia - this is very much a strategic-looking recommendation, whereas, conversely, recommendation 2, related to management of the CP, has primarily programmatic elements (such as revision of the M&E system). Finally, the recommendations do not clearly identify the responsible units for their implementation, per the evaluation guidance and the criterion. The recommendations vaguely reference the UNFPA CO and UNFPA APRO as target audiences for the recommendations. This is a weakness as it does not help to draft management responses, hence a 'partial' rating.
	iii Process for developing the recommendations is described, and includes the involvement of key stakeholders (e.g. evaluation reference group members), including those who will be affected by the recommendations.	Yes	The process for developing the recommendations is described in the preamble to the section, and it explicitly includes the involvement of key stakeholders. The evaluators clearly state that the recommendations are based on both the evaluation findings/conclusions and 'feedback received from key stakeholders'. The process included solicitation of feedback on preliminary findings from members of the ERG, the UNFPA AP Regional Office and the UNFPA Malaysia CO at the end of the field phase and in principle validation. Further, Chapter 1 (methodology) describes stakeholder involvement in the entire evaluation process, including in shaping the recommendations. The evaluation oversight process (i.e. the ERG) include the perspectives of stakeholders who work directly with beneficiaries and vulnerable groups and thus aimed to ensure their views were considered.
	iv Recommendations are clearly articulated and prioritized based on their importance, urgency, and potential impact.	Yes	The recommendations are assigned priority levels partially according to the guidance set by UNFPA - each of the recommendations is either "Very high" or "High" priority - there are no medium or low priority categorisations. This is mostly semantic, so the rating has not been affected. Further, the recommendations within each categorisation (strategic, programmatic, operational) are ordered from highest-priority to lowest, which is a good practice.
SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)		83%	Comments on Rating
Question 17. Does the evaluation report include all required information?			
	i Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, 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.	Yes	The opening pages of the report include all of the mandated/essential introductory elements such as the evaluation's name and timeframe, the report's date, the location of the evaluand (UNFPA Malaysia), details about the evaluation team and the commissioning organization, a list of the ERG members, acknowledgements, a comprehensive table of contents with listings for tables, figures, graphs, and annexes, and a list of acronyms and abbreviations. Additionally, the report features a 'key facts' table providing a snapshot of relevant country data (including sources), which is very useful.
	ii Annexes include, if not in body of report: terms of reference, evaluation matrix, list of respondents, results chain/ToC/logical framework, 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 (e.g. inception report), case study reports.	Yes	All of the required annexes are present, including a stakeholder map (annex 3) which provides a useful list of individuals targeted for interview by the evaluation - this can be compared with annex 4, which lists the ultimate interviewees (with personal names redacted for confidentiality, which is line with the UNFPA data protection policy). A minor issue is the use of a scanned version of the TOR for Annex 8 - the scanning process results in a file that is 20Mb+, which is impractical for publishing online. The original pdf text version of the TOR should have been located and used.
Question 18. Is the report logically structured and of reasonable length?			
	i The report has a logical structure that is easy to identify and navigate (for instance, with numbered sections, clear titles, well formatted).	Yes	The report is fully in accordance with the mandated structure for evaluation reports as laid out in the UNFPA Evaluation Handbook.
	ii Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR. <i>Note: Maximum pages for the main report, excluding executive summary and annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations and 50 for other types of evaluations</i>	No	The report is 90 pages long, excluding the executive summary and annexes, which is ten pages over the mandated limit (per the UNFPA Evaluation Handbook and the TOR) and 20 pages over the limit as per this criterion. The evaluation analytical framework (which is extensive) implies a very substantial amount of analysis to answer each of the evaluation questions with sufficient recourse to evidence, so the length is understandable, if not conforming to the UNFPA specifications. This said, for some of the preamble chapters (i.e. background) and many of the findings, there is scope for reducing the level of detail. As noted above, excerpting much of the reconciliation of the CPD results in the Effectiveness section (EQ5), specifically Table 4.1, would have reduced the report by several pages, and, similarly some of the unsubstantiated sub-findings the more descriptive text that simply lists activities could have been eliminated without detracting from the report - and indeed would have enhanced the remaining - better-quality - analysis.
Question 19. Is the report well presented?			
	i Report is easy to understand (written in an accessible way for the intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	Overall, the report reads well, with good use of language, terms and minimal use of jargon (instances where they are used, such as B40 or M40, are explained via footnotes). There are some, but not many, errors in writing or formulation of sentences. While the resources available for a comprehensive edit at CPE level may not have been to the same extent as a centralised evaluation, some of the errors (particularly when found in sections such as the conclusions or recommendations - which are more likely to be read) should or could have been noted by a relatively brief proof-reading. There are a range of relatively minor editing errors or issues that are not in conformance with the UNFPA style guide, for example: - use of acronyms on the cover/title page - mix of upper and lower case Capitalization for various terms (e.g. Evaluation Team) - page number and footer on the cover/title page While these do not affect the rating for this criterion, many of the issues have been noted for reference via in-line commentary.
	ii 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.	Yes	Although as noted above, at least one of the tables in the report could be removed to the annexes to reduce the overall page length, the report makes good use of a selection of tables and figures/charts to illustrate key analyses. All of these are labelled clearly and reflected in the list of tables/figures. Some of the graphics are not fully clear in terms of resolution or text size (e.g. figure 2.2) or have been stretched to fit the page (e.g. figure 2.3), or both (e.g. figure 2.4), but the details can be discerned. Given the nature of the report (i.e. not to be submitted for professional editing/publication) this is acceptable and does not affect the rating.
SECTION I: CROSS-CUTTING ISSUES (weight 10%)		69%	Comments on Rating
Question 20. Are cross cutting issues - in particular, human rights-based approach, gender equality, disability inclusion, LNOB - integrated in the core elements of the evaluation (e.g. evaluation design, methodology, findings, conclusions and recommendations)?			

i	Evaluation's data collection methods designed to capture the voices/perspectives of a wide range of stakeholders including right holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Partially	<p>Although the evaluators initially proposed engagement with rights-holders at community level, UNFPA Malaysia's work does not lead to direct engagement in on-the-ground activities and thus the direct link between UNFPA and rights-holders was stated not to exist. Therefore the evaluation focused on institutional stakeholders and rights-holder representative organisations. While the reasoning for this limitation is clear, and the list of UNFPA-supported interventions (provided in Annex 7) demonstrates a relatively limited number of these that might reasonably be directed at rights-holders, this modality of programming pertains in many UNFPA country programmes - only in some circumstances (such as humanitarian programming) does UNFPA engage in direct implementation.</p> <p>Thus, the rationale for absence of rights-holder voices is weakened, though there is still a partial justification given that the majority of programming is not rights-holder-focused.</p> <p>With respect to the logistics, the evaluators relied on online interviews, which is an acceptable substitute, though does limit the depth of interaction. Despite these limitations, the design of the data collection methods did attempt to be inclusive by consulting with a wide array of stakeholders, including organizations working with and representing marginalized populations. The interview/FGD tools (annex 6) do include a question specifically related to vulnerable groups, though it is somewhat generic (i.e. it does not specifically identify vulnerable groups such as PWDs).</p>
ii	Evaluation questions address cross cutting issues, such as human rights-based approach, gender equality, disability inclusion, LNOB, social and environmental standards as appropriate.	Partially	Three of the eight evaluation questions explicitly reference vulnerable or marginalised groups in their formulation, although LNOB is not referenced directly in the questions. A fourth question (#8) does refer specifically to "women, adolescents and youth" in the context as rights-holders, but does not do so explicitly in the context of vulnerability, hence the rating.
iii	Data is disaggregated by population groups (e.g. persons with disability, age, gender, etc.) where there are implications related to UNFPA's portfolio/interventions for these population groups; differential results are assessed (distribution of results across different groups).	Partially	<p>The methods section (1.3.2, para 19, page 8) notes that the evaluators sought "Disaggregated data (by, inter alia, gender, age, location, ethnicity, disability status and other relevant vulnerability criteria)". However, the nature of the work undertaken by UNFPA, and hence the focus of the evaluation, did not lend itself to disaggregation. For example, none of the CPD indicators presented in the result framework reconciliation under Effectiveness (EQs, table 4.1) have population numbers - all targets are institutional - and hence disaggregation is not feasible. While many of the initiative supported by UNFPA had a strong gender dimension (and are documented in the report), the reach of UNFPA-supported interventions to population level was not assessed and appeared to be of limited feasibility. The evaluators do note in the findings and recommendations that this lack of more granular outcome-related data is an issue that should be addressed.</p> <p>The evaluation does provide a gender-disaggregated list of respondents to the research in the methods description (table 1.2, page 12), which is positive.</p>
iv	Intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination (and how they overlap with each other) and how this may impact the performance or results of the evaluand.	Yes	<p>As discussed above, this CPE is atypical insofar as the subject CPD is more focused on organisational and institutional activities and outputs and less so (but still to a degree) on specific impacts on population groups. Thus, the ability of the evaluators to disaggregate data according to intersecting vulnerabilities is significantly constrained - and more so by the limited quality of the secondary data available to the evaluators (noted as a challenge for UNFPA to address).</p> <p>The issue of data disaggregation by intersecting vulnerabilities aside, the report does seek to evaluate the quality of UNFPA's work to reach groups susceptible to these - conclusion 6 and related recommendation 6 explicitly note this and encourage UNFPA to more deeply integrate these considerations into future programming. Therefore, it appears that the evaluators have made some efforts to ensure that their findings and conclusions/recommendations reflect an intersectional analysis, within the limitations of the overall evaluation scope and methods.</p>
v	Findings, conclusions and recommendations, address cross-cutting issues such as equality and vulnerability, disability inclusion, leave no-one behind, social and environmental as relevant.	Yes	<p>The evaluation report's findings indicate an attempt to address the relevant cross-cutting issues in the analysis. The report analyses the integration of gender equality and disability inclusion across different programmatic areas, noting areas of progress and shortcomings, such as the limited focus on intersectionality and the lack of a dedicated monitoring framework for gender and social inclusion. While the cross-cutting issues are an integral part of the analysis on vulnerable populations and SRHR, the report also introduces environmental considerations, particularly climate change, as key emerging megatrends that require integration into future programming.</p> <p>The conclusions demonstrate a similar engagement with cross-cutting issues. For example, Conclusion 6 directly addresses the extent to which diverse and marginalized populations, including PWDs, were considered, highlighting limitations in the LNOB principle and in addressing intersectionality and inclusivity for the most vulnerable. Regarding social and environmental aspects, Conclusion 5 explicitly notes a need to holistically account for "climate change challenges" and their intersection with demographic megatrends.</p> <p>The recommendations are formulated to address these findings by, for example suggesting more targeted partnerships, enhanced data collection and analysis, and a more holistic TOC for the next Country Programme.</p>
vi	Inclusion of young people in the evaluation team and/or Reference Group [N/A if not requested in ToR]	Yes	The cover pages list the evaluation team and include a young and emerging evaluator.
Question 21. Does the evaluation meet UN SWAP evaluation performance indicators?		5	Comments on Rating
<i>Note: this question will be rated according to UN SWAP standards with detail provided below.</i>			
i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Satisfactorily integrated	<p>The evaluation explicitly refers to a "Gender, Human Rights-Based and LNOB approach" (noted in the executive summary and the Evaluation Approach, section 1.3). The scope includes the programmatic area of Gender Equality and Women's Empowerment (GEWE), and is explicitly focused on assessing the country programme output area of gender and social norms change. Furthermore, the methodology highlights throughout the integration of gender principles and addressing the needs of marginalized populations in evaluation questions.</p> <p>While the evaluation purpose and objectives are not focused on GEEW, the evaluation criteria and questions are well-designed to ensure the collection of GEEW-related data. Table 1.1 outlines the Evaluation Criteria and Evaluation Questions, with EQ1 (Relevance) asking about the extent to which the country programme is adapted to the needs of diverse populations, including vulnerable and marginalized groups such as women. More specifically, EQ5 (Effectiveness) directly inquires about the "advancement of gender equality and the empowerment of all women and girls" as a key area to assess the delivery of intended results.</p>
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Partially integrated	<p>While the design of the evaluation emphasised inclusion of a gender, human rights-based and LNOB approach, and noted an intention to use disaggregated data (by, "inter alia, gender, age, location, ethnicity, disability status"), the actuality of GEWE considerations in the design is more limited.</p> <p>The sampling frame aimed to address the diversity of stakeholders affected by the intervention, focusing on major categories across the Country Programme's themes and outputs, and specifically noted inclusion of stakeholders who work directly with the beneficiaries/vulnerable groups that are furthest behind, such as local organizations, NGOs, and community leaders - this is evident in the selection of participating organisations - e.g. 4-5 of the participating NGOs/CSOs are specifically focused on women's rights or similar. As discussed above, participation of rights holders directly in the evaluation was not undertaken for various reasons, but the inclusion of organisations that represent them was a positive step.</p> <p>The design of tools was less comprehensive in the integration of these issues - as noted above, the interview guides had a single question (albeit in multiple parts) related to vulnerable groups - there is no specific mention of GEWE or related considerations. Finally, quantitative data (such as was available or analysed) was not disaggregated by gender.</p>
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	<p>The evaluation does consider issues of gender and human rights (including intersectionality) across the findings/analysis, with these are well reflected in the conclusions section. Specifically, Conclusion 6 directly addresses consideration of diverse and marginalized populations, including women, in programme design and SRHR/GBV services. Conclusion 8 explicitly notes the challenges faced by interventions in adopting gender-transformative approaches that challenge underlying norms, a critical analytical gap also identified in the findings alongside the absence of a dedicated gender and social inclusion monitoring framework.</p> <p>The evaluation also includes specific recommendations that integrate gender considerations (#s 3, 4, 6, 7), including priorities for action to improve gender equality of the future country programme. Intersectionality is not covered in the background nor is data disaggregated or the voices of different groups present directly in the report. Hence the criterion is satisfactorily but not fully met.</p>

List of SDGs

1. No Poverty
2. Zero Hunger
3. Good Health and Well-being

Three transformative results

1. Ending unmet need for family planning
2. Ending preventable maternal deaths
3. Ending gender-based violence and harmful practices

4. Quality Education
5. Gender Equality
6. Clean Water and Sanitation
7. Affordable and Clean Energy
8. Decent Work and Economic Growth
9. Industry, Innovation and Infrastructure
10. Reduced Inequality
11. Sustainable Cities and Communities
12. Responsible Consumption and Production
13. Climate Action
14. Life Below Water
15. Life on Land
16. Peace, Justice and Strong Institutions
17. Partnerships for the Goals

Six outputs
1. Policy and accountability
2. Quality of care and services
3. Gender and social norms
4. Population change and data
5. Humanitarian action
6. Adolescents and youth
Six accelerators
1. Human rights-based and gender-transformative approaches
2. Innovation and digitalization
3. Partnerships, South-South and triangular cooperation, and financing
4. Data and evidence
5. Leaving no one behind and reaching the furthest behind first