Evaluation Branch

Division for Oversight Services (DOS)

QUALITY ASSESSMENT OF UNFPA DECENTRALISED COUNTRY PROGRAMME EVALUATIONS

17 February 2012
New York

UNFPA Evaluation Quality Assessment Managers

Louis Charpentier
Chief, Evaluation Branch, DOS

Hicham Daoudi
Evaluation Adviser, Evaluation Branch, DOS

Assisted by

Olivia Roberts
Evaluation Analyst, Evaluation Branch, DOS

Contents

Diagrams

Tables

Acronyms

EQA	Evaluation Quality Assessment
CPE	Country Programme Evaluation
CO	Country office
CP	Country programme
RO	Regional Office
M&E	Monitoring and Evaluation
ToRs	Terms of Reference

Introduction

As per the UNFPA evaluation policy,[1] the Evaluation Branch at DOS is responsible for providing assurance on the quality of evaluations. This report presents the results of the evaluation quality assessment (EQA) completed by the Evaluation Branch in 2012, and covering evaluations conducted in 2010 and 2011.

Following a request from the Executive Board in 2009,[2] that UNFPA programmes be evaluated at least once during their cycle, all programmes due to end in 2010-2011 were subject to an evaluation prior to their renewal. The country programme evaluation (CPE) coverage thus rose from 14% in 2009 to 80% in 2010,[3] and to 100% in 2011. In order to assess whether this significant quantitative progress was accompanied by a parallel improvement in the quality of CPEs, the Evaluation Branch decided to fully devote its 2012 EQA exercise to decentralized CPEs.

In addition to this quality assurance objective, the decision of the Evaluation Branch to focus its 2012 EQA on CPEs was also guided by its willingness to gain a better understanding of the gaps and challenges that country offices (COs) face in the design and implementation of CPEs. This, in turn, informed the production of a CPE methodology,[4] tested by the Evaluation Branch in the two pilot evaluations of the Cameroon and the Bolivia country programmes. The activities of the Evaluation Branch in: (i) producing a CPE methodology, (ii) conducting CPEs and (iii) ensuring quality assurance do indeed complement one another with a view to increasing the quality, and hence the use, of evaluation products in UNFPA.

The EQA process uses eight specific assessment criteria to measure the quality of evaluation reports: i) Structure and Clarity of Reporting; ii) Executive Summary; iii) Design and Methodology; iv) Reliability of Data; v) Findings and Analysis; vi) Conclusions; vii) Recommendations; and viii) Meeting Needs. Each CPE report is assessed against all eight criteria, and the results are collated to provide an overall assessment of the report on a four-level rating scale: *Unsatisfactory*, *Poor*, *Good* or *Very good*.

Contents of the report

The first part of the report presents the results of the 2012 EQA performed by the Evaluation Branch.

In a second part, the report presents the results of a survey of all COs that conducted a CPE in 2010 and 2011. The online "CPE survey" included questions on: the planning and management of CPEs as well as the resources committed to CPEs by country offices.

The third part of the report consists of a set of overall conclusions and recommendations derived from both exercises.

---

[1] UNFPA Evaluation Policy, 2009.
[2] Decision 2009/18 8f, Executive Board, 2009.
[3] DP/FPA/2011/3 (part 1) para 102.
[4] Evaluation Branch Handbook on How to Design and Conduct a Country Programme Evaluation, 2012.

Comparison with previous reviews of evaluation quality

Previous reviews of the quality of evaluations at UNFPA have been conducted, along with a meta-evaluation in 2005[5] and the 2009 EQA.[6] These assessments utilised methodologies that differ both from each other and from the 2012 EQA. It is therefore not possible to directly compare the 2012 EQA results with those of the 2005 meta-evaluation and of the 2009 EQA. Instead, this report will, where relevant, reflect on the quantitative and qualitative results from both these reports in the context of the specific findings of the 2012 EQA and identify any similarities or changes (positive or negative) only where legitimate comparisons are possible.

---

[5] Strengthening Evaluation for Improved Programming: UNFPA Evaluation Quality Assessment, 2005.
[6] Evaluation Quality Assessment: 2007-2008, 2009.

Part I. Quality Assessment of the Decentralized Country Programme Evaluations Conducted in 2010 and 2011

The Evaluation Branch performs an Evaluation Quality Assessment (EQA) on a biennial basis to measure the quality of decentralised evaluations in UNFPA. The EQA process aims to:

i) provide constructive feedback on the quality of individual evaluation reports;

ii) provide an independent assessment of the quality and usefulness of reports for senior managers (headquarters, regional offices, country offices) and the Executive Board ;

iii) measure and analyse trends across the organisation;

iv) provide a set of actionable recommendations with a view to improving the quality of evaluation reports.

1.1 Evaluation Quality Assessment Process

In 2011, the Evaluation Branch introduced a new EQA tool (the EQA grid)[7] in order to bring UNFPA practice in line with the standards applied by other international organizations and, in particular, by other UN agencies.[8] When designing the grid, the Evaluation Branch also took account of the need to ensure that the EQA grids sent to country offices (COs) are user-friendly and clearly convey all key findings and advisory comments.

As of May 2011, the EQA grid and explanatory note must be annexed to all Terms of Reference (ToRs) for on-going and future evaluations commissioned by UNFPA.[9] This was a recommendation from the 2005 UNFPA meta-evaluation which stated that "the EQA matrix should be adapted to a set of quality standards that could be attached to the evaluations' ToRs". Furthermore, monitoring and evaluation (M&E) officers/focal points as well as other staff managing evaluations in COs must also use the EQA grid to assess the quality of the reports.

As previously stated, COs are strongly encouraged to familiarise themselves with the EQA grid to gain a better understanding of all the necessary components of a good quality evaluation report. To this end, the Evaluation Branch organized several trainings on the EQA in 2011, which were attended by all regional office M&E advisers and some M&E officers (including most M&E officers of the African region).[10]

---

[7] EQA grid and accompanying explanatory note; see Annex 1.

[8] http://www.unicef.org/evaldatabase/files/UNICEF_Eval_Report_Standards.pdf

[9] Email from Dr. Olivier Brasseur, Director of DOS to all UNFPA Representatives, 12 May 2011.

[10] Including a training as part of the Results Based Management: RBM Cluster Meeting Central Africa, November 2011, Yaoundé, Cameroon. For further information see http://www.unfpa.org/public/home/about/Evaluation/Methodology

The EQA is based on eight assessment criteria:

1. **Structure and Clarity of Reporting** - To ensure the report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.
2. **Executive Summary** - To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation.
3. **Design and Methodology** - To provide a clear description and rationale for the methodology used, including key processes and limitations
4. **Reliability of Data** - To clarify data collection processes and data quality
5. **Findings and Analysis** - To ensure sound analysis and credible findings, based on evidence and with clear cause and effect links
6. **Conclusions** - To ensure conclusions are a reasonable judgment stemming from the findings
7. **Recommendations** - To assess the usefulness of recommendations
8. **Meeting Needs** - To ensure that the evaluation report responds to the requirements stated in the ToR (the ToR must be annexed to the report)

When reviewing an evaluation report, a rating of *Unsatisfactory*, *Poor*, *Good*, or *Very good* is given for each criterion.

Diagram 1. EQA Scoring matrix

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Unsatisfactory	Poor	Good	Very good
5. Findings and analysis (50)		50		
6. Conclusions (12)		12		
7. Recommendations (12)		12		
8. Meeting needs (12		12		
3. Design and methodology (5)		5		
4. Reliability of data (5)		5		
1. Structure and clarity of reporting (2)		2		
2. Executive summary (2)			2	
TOTAL		98	2	

Each EQA criterion is associated with a multiplying factor out of 100 which is proportionate to, and illustrates, its relative importance as regards the overall quality of the report. For example, *Findings and Analysis* is the most prominent of all 8 criteria and has a multiplying factor of 50. If this criteria is given an assessment level of *Poor* for an evaluation report, then 50 is placed in that column (see Diagram 1). The scores for each assessment level are then calculated, and the level that scores the highest number in total determines the overall rating for the report. In Diagram 1, as the highest score is in the *Poor* column (98), the report is rated as *Poor*.

The aim of the overall assessment level is not to grade or compare reports. Instead, it expresses an objective judgment on the quality of the entire evaluation report as well as on each evaluation criterion used. It also enables the measurement of progress (or lack thereof) over time, either in the quality of UNFPA funded evaluation reports or for each specific quality criterion.

Another key aspect of the EQA grid is the narrative that accompanies each criterion and which provides the reasoning for the assessment level. Annex 2 provides an example of a completed EQA grid, including a narrative.

1.2 Conduct of the EQA

As previously mentioned, and unlike previous quality assessments, it was decided to focus the 2012 EQA on decentralized CPEs. As a result, a total of 34 country programme evaluation reports were considered eligible for assessment[11] and were reviewed in accordance with the process described above. Evaluation Branch staff conducted the review, with each staff member assigned a number of CPE reports. Each EQA was then peer-reviewed by a senior member (evaluation adviser) of the Evaluation Branch to ensure consistency of the process and harmonization of the products. The finalized EQA grid was then uploaded, along with the original CPE report, to the publicly-available Evaluation Database.[12] The EQA was subsequently sent by the Chief of the Evaluation Branch to the Representative of the relevant UNFPA country office with information on the process and an explanation of the overall assessment of the evaluation report. In its communication with COs, the Evaluation Branch strongly emphasized that the assessment level should be seen as an indicator of the CO current methodological challenges in the management of CPEs. Indeed, instead of a judgement on the CO performance, the EQA should be viewed as a starting point from which improvement in the quality of evaluations should be measured.

1.3 Main Results

Thirty-four country programme evaluations were included in the 2012 EQA, with reports received from all regions (see Diagram 2). Table 1 (below) contains the results of the 2012 EQA by assessment level, in terms of the number of reports achieving each assessment level and as a percentage of the total number of reports reviewed.

Table 1. 2012 EQA Results

Assessment level	Number of reports	As percentage of total reports
Very good	0	0%
Good	3	8.8%
Poor	23	67.6%
Unsatisfactory	8	23.6%

---

[11] Several country offices conducted their CPE either before the period specified in the Evaluation Guidelines or after the designated completion date, and were therefore excluded from this review.

[12] http://web2.unfpa.org/public/about/oversight/evaluations/. Annex 4 contains a summary of all 34 EQAs.

Overall, more than 90% of CPE reports were rated as *Poor* or *Unsatisfactory* (31 out of 34 reports). Three reports were rated as *Good*, with no CPE reports rated as *Very good*.

These results show that the quality of CPE reports in UNFPA remains a concern and are indicative of the fact that there seems to have been no improvement in evaluation quality from the previous assessments in 2005 and 2009.[13]

1.4 Results by region

Of the 34 CPE reports received for review in 2011, almost 40% were received from UNFPA country offices in Africa.

Diagram 2. CPE reports by region

---
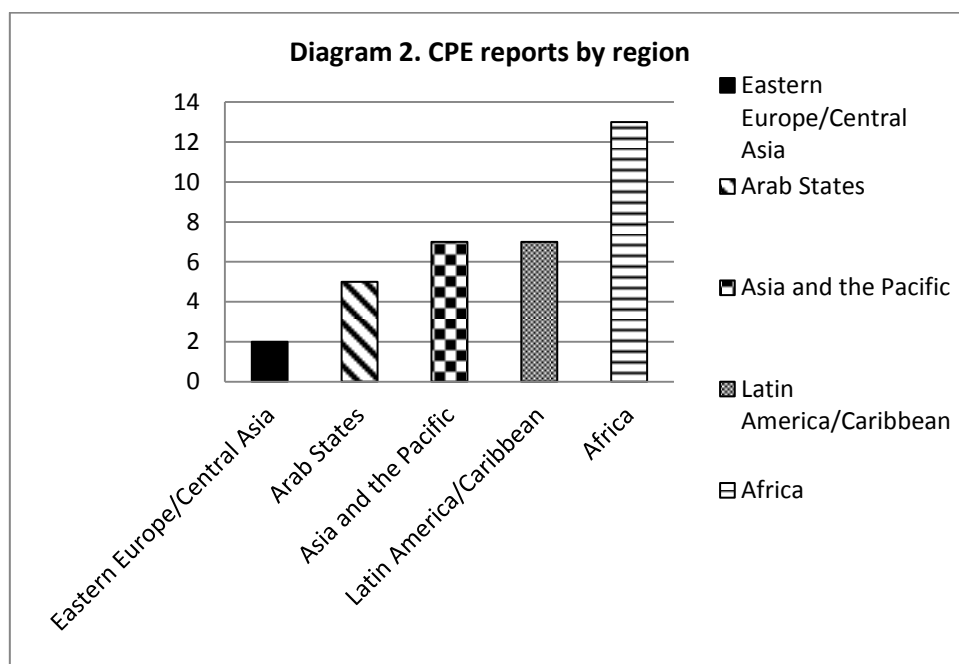
[13] The 2005 meta-evaluation assessed 66% of evaluation reports as being unsatisfactory across all criteria and stated that "evaluation quality at UNFPA is currently unsatisfactory and requires improvement" (2005 meta-evaluation, p.9.). The 2009 EQA report stated that "no systematic improvement in evaluation quality in the period 2007-2009 was identified" and that 51% of evaluation reports were below expectations of quality (2009 EQA, p.21.).

Diagram 3. 2012 EQA results by region

Region	Good	Poor	Unsatisfactory
EECA		2	
Africa	1	6	6
Arab States		4	1
LAC		6	1
Asia Pacific	2	5	
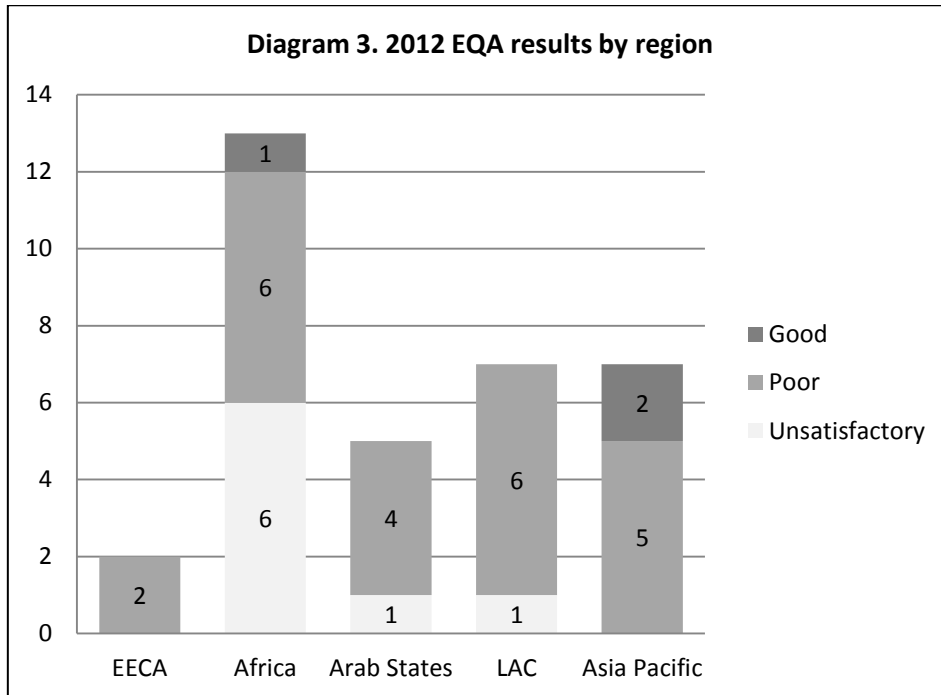
The 2005 meta-evaluation stated that weak evaluation performance was an issue found across the organisation; this is confirmed by the results of the 2012 EQA, with only 2 regions having a CPE report that was assessed as *Good*.

1.5 Results by Assessment Criteria

Sections 1.5.1 to 1.5.8 present an analysis of the results for each assessment criterion. Examples from individual EQAs are provided in text boxes.

With over 90% of reports rated *Poor* or *Unsatisfactory, Findings & Analysis* as well as *Conclusions* are the main areas of weaknesses in CPE reports. The strongest criteria across all reports were *Structure & Clarity of Reporting* and *Executive Summary*, with almost 30% of reports rated as *Good* for both criteria. The results also pointed at weaknesses in terms of *Methodology* and *Reliability of Data,* and in the formulation of prioritised and actionable *Recommendations*.

1.5.1 Structure and Clarity of Reporting

This criterion assesses whether the report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.

The structure of a CPE report is extremely important to ensure that the reader gains a clear understanding of the objectives and the results of the evaluation. The structure of most CPE reports was consistent with recognized international good practice.[14] This was the highest rated criteria overall, with almost a third of CPE reports assessed as *Good*.

However, some evaluators who conducted the CPEs chose to **change the content and sequence of sections within the CPE reports**, for example combining sections (most often, Findings and Recommendations). These modifications are inappropriate as they make it extremely difficult for the reader to follow the logical flow from findings to conclusions and to recommendations.

EQA Brazil: "Recommendations addressing focus areas are presented at the end of each [Findings] section while programmatic recommendations are presented in the final chapter, which dilutes the message. This structure is not ideal and puts the clarity of the report at risk."

A common flaw in the CPE reports was that the Conclusions chapter was often located incorrectly. For example it can be found after the Recommendations or is combined with Lessons Learned. At times, the Conclusions are missing entirely. This modification again interrupts the logical flow for the reader and compromises the usefulness of the report.

EQA Eritrea: "There is no actual section on conclusions; although it is entitled conclusions it does not provide the reader with evaluation conclusions."

Some CPE reports are **extremely long**, particularly key sections such as Findings. Overly-long sections undermine the content of the report as they are indicative of a lack of focus.

EQA Bangladesh: "Overall there are some interesting evidence-based insights but these are lost in the volume of unreferenced description and unhelpful structure of this section."

Annexes were correctly used in the CPE reports to provide relevant supplementary information, such as methodological tools (interview guides, list of documents consulted). The use of annexes can help maintain brevity in the report whilst ensuring that key information is provided, such as evidence to substantiate arguments. However, some CPE reports used annexes to include information that had no clear link to the rest of the report, or the relevance of which was not explained.

---

[14] As detailed in the EQA grid (Annex 3), the suggested CPE report structure is: i) Acronyms; ii) Executive Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned (where applicable). Minimum requirements for the Annexes are: ToRs; Bibliography; List of interviewees; Methodological instruments used.

1.5.2 Executive Summary

This criterion assesses whether the executive summary reads as a standalone section and is a useful resource in its own right, presenting main results and key aspects of methodology, conclusions and recommendations.

All CPE reports included an executive summary. However, many of them were **extremely long** - up to 15 pages - whereas a *Good* executive summary should not exceed 3 to 4 pages. In spite of their length, many executive summaries lacked key information. Some failed to refer to the purpose and objectives of the evaluation or to provide an overview of the evaluation results. Such gaps affect the ability of the report to convey key messages to the readers. Most summaries in the reports reviewed **did not function as a stand-alone section**. Only a quarter of reports included a summary section that was 'fit for purpose' and were consequently rated as *Good.*

> EQA Sao Tome: "The executive summary contained little apart from a list of 25 findings, with a single paragraph outlining the purpose of the evaluation, and no mention of methodology or design."
>
> EQA Honduras: "The executive summary is not a stand-alone section and does not provide a complete overview of the evaluation. There is no mention of the purpose, intended audience, and objectives of the evaluation. The programme is not described and there is no mention of the methodology. Findings and conclusions are presented but recommendations are missing."

Although they remain unsatisfactory, these results mark an improvement from the 2005 meta-evaluation and the 2009 EQA. These quality assessments reported that 45% and 19% of evaluation reports were missing executive summaries respectively.[15]

1.5.3 Design and Methodology

This criterion assesses whether the methodology is clearly explained, methodological choices are justified and limitations are made explicit.

Three-quarters of CPE reports were rated as *Poor* or *Unsatisfactory* against this criterion. This is a decrease in quality from the 2009 EQA which found that over 50% of evaluation reports were based on methodologies "sufficiently weak to call into question the validity of any findings and thus, conclusions and recommendations".[16]

The 2012 EQA shows that there was inadequate discussion of methodological choices in the CPE reports, and the explanation for the choice of an evaluation method over alternatives was often insufficient. **Methodologies were often poorly explained**, and there were significant gaps in the provision of key information, such as justification of tools selection. Methodological instruments were mostly located in the annexes of CPE reports, but in some cases were missing entirely, or lacked sufficient detail.

---

[15] 2005 meta-evaluation, p.10.; 2009 EQA, p.20.
[16] 2009 EQA, p.14.

While *Good* methodological sections should recognise and discuss limitations to the effective conduct of an evaluation, many CPE reports under review only mentioned that some challenges were encountered and **frequently failed to discuss the steps taken to overcome these constraints**. Furthermore, the confusion was often made between methodological and logistical limitations (transportation issues, insufficient time for the evaluation in the ToRs, failure to translate key documents).

EQA Ghana: "Little information is provided: the evaluators 'developed and applied a set of tools', with no further details provided. There is no detailed list of stakeholders, nor how they were contacted, with only reference to 'program implementers'. No questionnaire is provided in the report. There is no discussion of data collection, triangulation or methodological choices."

However, it must be stressed that in some instances, reports contained clear and detailed methodological description and indicated how data challenges were addressed.

1.5.4 Reliability of Data

This criterion assesses both whether the data used is credible and well-referenced, and whether attempts were made to identify alternative sources of information when data availability was limited.

CPE reports frequently mentioned the **absence of key data** necessary for the evaluation, particularly baseline data. Systemic weaknesses in the design of country programmes results frameworks partly account for this. However, CPE reports were rated as *Poor* when no appropriate action was taken to address this challenge, such as identifying and using alternative data sources.[17]

EQA Senegal: "The report contains little data and in particular lacks data regarding the outputs, results and impacts of the programme. Although the evaluators rightly point out the absence of baseline data, there is no evidence of their attempt to establish or reconstruct these data."

References were often either **incomplete or missing entirely.**

EQA Chad: "It is not clear what data supports all the activities-related findings presented in the results section of the report. Quantitative data are never referenced, nor are any interviewees and field visits mentioned in the text or a footnote. Some very basic indicators are never provided in the report, not least the Maternal Mortality Ratio."

This criterion was not used in previous EQAs, thus preventing comparison with the results of the 2012 EQA.

---

[17] Such as relevant reports from other UN agencies.

1.5.5 Findings and Analysis

This criterion assesses whether findings are based on evidence and stem from a sound and rigorous analysis.

Over 90% of CPE reports were rated *Poor* or *Unsatisfactory* for the quality of their findings and analysis, which is the second lowest rated assessment criterion overall. The reviewed CPE reports frequently lacked a sound and rigorous analytical work.

In most CPE reports, **the UNFPA contribution to development results is not clearly identified.** Moreover, when results are reported, their causal link with UNFPA interventions are generally not discussed.

EQA English and Dutch speaking Caribbean: "No link has been made between activities, outputs and achievement of results. Achievements attributed to programme outputs are described in general terms ("increase in CPR") rather than more specific indicators."

The discussion of evaluation findings often remains at the level of completion of activities, resulting (at best) in documents akin to monitoring reports. Furthermore, rather lengthy descriptions of activities sometimes contribute to the generally excessive length of evaluation reports highlighted earlier. [18] These detailed sections often lack any critical discussion of the activities.

EQA Chad: "The report's findings and analysis section is entirely confined to an assessment of activities and does not present any discussion of progress towards CPAP results."

EQA Eritrea: "No analytical work can be found in the report, which remains very descriptive. "Findings" at best consist in factual observations, most often anecdotal, without any visible effort to test cause and effect assumptions."

Analysis in the CPE reports is also, at times, overly positive about the UNFPA contribution, without any substantiating evidence.

EQA Kyrgyzstan: "The CO activities are wholly described in positive terms e.g.; 'UNFPA is deeply involved …'; 'UNFPA has undeniably contributed to the enhancement of the reproductive health status of the people of [xxx]…'; 'UNFPA has been actively involved in all these areas'. Overall, this reviewer could find few examples of criticism of the CP by the evaluator."

EQA Malawi: "In particular the section on country programme impact affirms UNFPA contribution to a number of indicators without providing any evidence of that contribution e.g.; 'it is clear from all IPs, communities and key informants that the 6th CP has made tremendous strides in contributing to the development goals and objectives of Malawi, hence contributing to the quality of life of her people', 'CP has made enormous contributions at all levels in the country'."

---

[18] see 1.5.1 Structure and Clarity of Reporting.

With regard to the *Findings and Analysis* criterion, the 2012 EQA confirms the results of the 2009 EQA which reported that "the most common deficiency in evaluation reports was that the statement of findings was unsupported by data."[19] The 2005 meta-evaluation stated similar findings: "[UNFPA] evaluations are stronger in drawing inferences and broad conclusions rather than [...] establishing findings."[20]

1.5.6 Conclusions

This criterion assesses whether the conclusions are based on credible findings and convey the evaluators' unbiased judgment on the intervention.

The formulation of conclusions is an important process through which evaluators express their judgment on the intervention and identify the key issues to be addressed in the recommendations section. The 2012 EQA reveals a **systematic lack of understanding of the function of the conclusions section in CPE reports**. The Conclusions criterion is the lowest rated of all assessment criteria, with almost half of the reports rated *Unsatisfactory* and 90% rated either *Poor or Unsatisfactory*.

> EQA Sao Tome: "The Conclusions section is less than a page, and makes very generic statements (e.g.; 'there have been changes in indicators at both outcome and output levels for all three programme components'). This section performs little useful function in terms of evaluation".

The Conclusions section was frequently missing or ill-structured. Many CPE reports included a very limited section called Conclusions that simply listed various 'strengths and weaknesses' of the CP, or contained general statements. Conclusions also frequently lacked a clear link to the Findings and also often constituted one of the shortest chapters in CPE reports.

These criticisms are significant as the quality of this section affects the rest of the report. *Poor* or *Unsatisfactory* conclusions limit the communication of the evaluators' judgment on interventions. The lack of clear conclusions also hampers the development of useful recommendations.

> EQA Thailand: "The detailed findings have not been translated into effective conclusions, and it is difficult to distinguish what were the achievements or shortcomings of the CP that were outlined in the Findings section."

In the 2005 meta-evaluation, 42% of Conclusions sections were *Missing* or deemed *Unsatisfactory*[21] and the results from the 2009 EQA did not indicate any improvement in quality for this criterion.[22]

---

[19] 2009 EQA, p.17.
[20] 2005 meta-evaluation, p.35.
[21] 2005 meta-evaluation, p.10.
[22] 2009 EQA, p.18.

1.5.7 Recommendations

This criterion assesses whether there is a logical flow from the conclusions to the recommendations, which should then be presented in priority order and made as operational as possible.

Over 80% of CPE reports were rated *Poor* or *Unsatisfactory* against this criterion. In particular, the following issues were repeatedly identified:

i) recommendations were **not prioritized**;

ii) they were insufficiently targeted and lacked clear deadlines, **which hampered their operational feasibility**;

iii) recommendations were often **too numerous** (56 recommendations in one CPE report, and another report included 34 recommendations for a single mandate area).

> EQA Peru: "The Recommendations are not clustered, too general in some cases and should be more strategic and targeted. They should also be presented in priority order, and the targeted audience and feasibility is not mentioned. Overall, there are too many recommendations."
>
> EQA Sao Tome: "Recommended Actions often lack detail (e.g.; 'build capacity for the collection and analysis of data at district level'), are impractical (e.g.; 'improve quality of care and services in all components of reproductive health'), or are simple statements of problems rather than being action statements."

The Recommendations section was often combined with other elements of the evaluation report, such as findings. As mentioned earlier, a poor structure has a negative impact on the usefulness of a report.[23] Recommendations should be found in a distinct section, after the conclusions section, while the logical flow leading successively from findings to conclusions then to recommendations should be made as clear as possible.

Recommendations that are poorly written and/or insufficiently linked to conclusions are unlikely to be used in decision-making processes by the CO, such as the development of the next programme cycle.

The results of the 2012 EQA with regard to the Recommendations criterion are a marked departure from previous quality assessments. The 2009 EQA judged that 70% of evaluation reports met or exceeded expectations for recommendations,[24] which was a slight decrease from the 78% of recommendations that were rated satisfactory or above in the 2005 meta-evaluation.[25]

---

[23] See section 1.5.1 Structure and Clarity of Reporting.
[24] 2009 EQA, p.19.
[25] 2005 meta-evaluation, p.10.

1.5.8 Meeting Needs

This criterion assesses whether the evaluation report adequately addresses the information needs and responds to the requirements stated in the terms of reference.

The terms of reference (ToRs) are of critical importance to the design and conduct of a CPE. They describe the intended scope and focus through the identification of the objectives and main evaluation questions. The ToRs also enable the evaluation manager to judge whether the final CPE report meets the needs of the CO. The 2005 meta-evaluation already called the attention of evaluation managers to the necessity to produce ToRs with a realistic and clearly delineated scope as a prerequisite to a good quality evaluation report.[26]

In several instances, **the CPE ToRs extended beyond normal expectations for a CPE**. For example some CPE ToRs included evaluation questions on issues that would be better suited to another type of review (such as questions on financial management). At design phase it is the responsibility of the evaluators to critically discuss the ToRs with a view to ensuring the feasibility of the evaluation. However, evaluators often did not, as would be expected, suggest modifications to the ToRs and/or propose to exclude unjustified demands from the evaluation scope.

CPE reports frequently mentioned that the **short duration for the evaluation stated in the ToRs was a limitation,** either because the ToRs were over-ambitious or time allocated for the conduct of the evaluation was insufficient. For example, one CPE report stated that "the scope of the evaluation is overly ambitious given the time period available. Time constraints limit meaningful measurement of achievements or behaviour change (e.g.; measure of capacity development)." The 2005 meta-evaluation also identified that "current time spent on evaluations is too short." [27]

EQA Chad [CPE duration six weeks]: "The report amounts more to some sort of monitoring rather than a fully-fledged CP evaluation which probably was not feasible in such a short period of time."

EQA Mauritania: "ToRs foresee only 25 work-days which is far from being sufficient to conduct the fully-fledged CPE as suggested here by ToR's nine specific objectives and 23 evaluation questions."

Finally, many CPE final reports were issued without the related ToRs in annex. This confirms a similar finding from previous EQAs; the 2005 meta-evaluation and the 2009 EQA already reported that ToRs were missing for 54% and 36% of reports respectively.[28]

---

[26] 2005 meta-evaluation, p.10.
[27] 2005 meta-evaluation, p.9.
[28] 2005 meta-evaluation, p.17.; 2009 EQA, p.8.

Part II. Country Programme Evaluation Survey

In May 2011, the Evaluation Branch at DOS led a survey covering the 40 country offices (COs) which were to conduct a CPE in 2010 and 2011.[29] The survey contained 41 questions grouped under three themes: (i) the planning, (ii) the management and (iii) resources (budget and staff) allocated by COs for CPEs. COs were asked to complete the survey within four weeks, and to answer all questions. A 100% response rate was achieved.

Staff from the Evaluation Branch analysed the responses from COs using both quantitative and qualitative analytical methods, and the results are presented in sections 2.1-2.3.

Diagram 4. Regional distribution of UNFPA country offices participating in the CPE Survey

- Eastern Europe/Central Asia
- Asia Pacific
- Africa
- Arab States
- Latin America/Caribbean
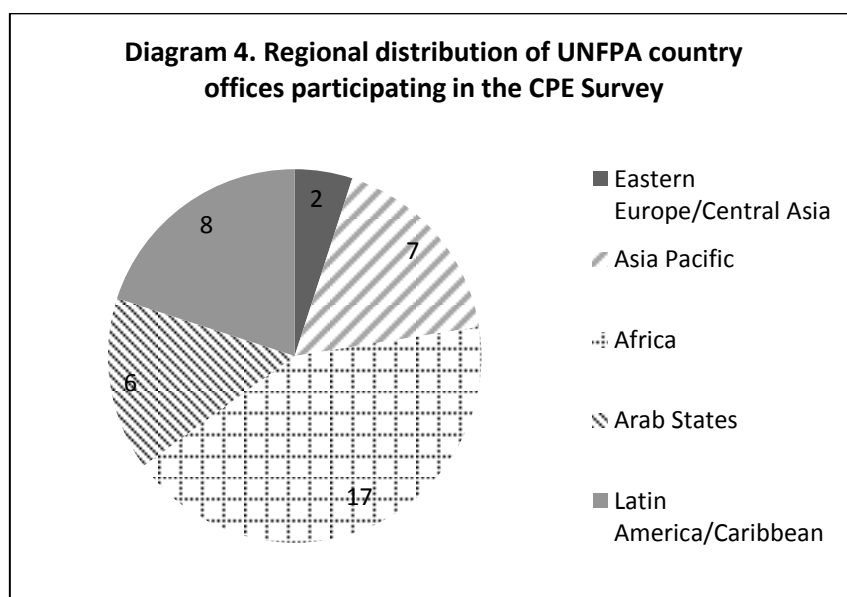
2.1 Planning of Country Programme Evaluations

This section provides an analysis of responses to CPE survey questions on: (i) the duration of the evaluation, (ii) whether the CPE was budgeted and (iii) its actual cost as compared to the cost and number of the other evaluations and reviews conducted during the programme cycle, including mid-term reviews.

2.1.1 Duration of the CPE

The average duration of an evaluation exercise was three months, and the reported range was from one to eight months. Three months is an extremely short period of time for an evaluation process that should encompass the following phases: (i) preparation; (ii) design; (iii) data collection and

---

[29] In line with the Executive Board request that UNFPA programmes be evaluated at least once during their cycle (Executive Board decision 2009/18, 8f).

analysis; and (iv) reporting. The results from the 2012 EQA identified the short time period specified in the Terms of Reference (ToRs) as a key limitation to the conduct of a good evaluation.[30] The 2005 UNFPA meta-evaluation also highlighted the lack of time allocated to evaluations (the average duration of evaluations in 2005 was 20 days), and concluded that "while evaluation quality depends on several factors, there appears to be a correlation between length of time spent on the evaluation and evaluation quality....[and] some country offices were not clear on what constitutes an optimal amount of time for an evaluation".[31]

Diagram 5. Proportion of CPEs budgeted at the beginning of the country programme
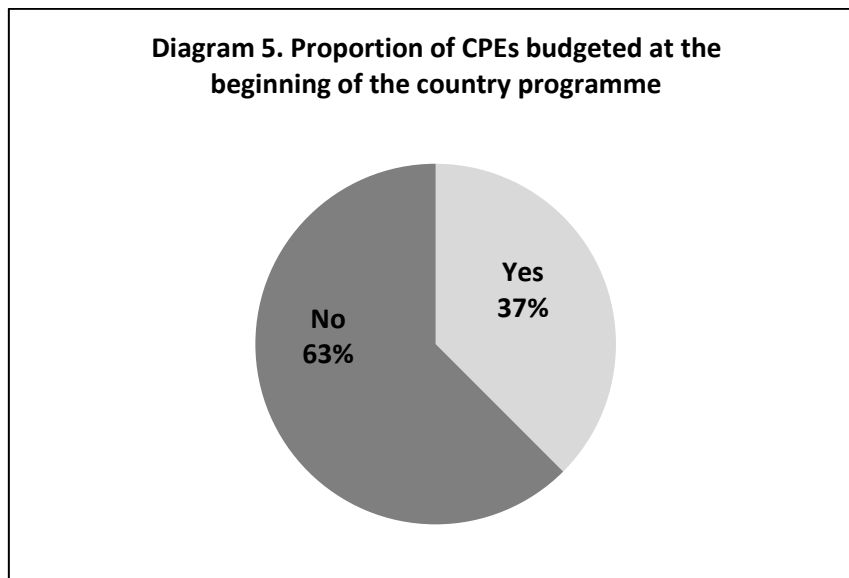
Almost two-thirds of CPEs were not budgeted at the beginning of the programme cycle**.** In Latin America and the Caribbean region, 75% of COs had not earmarked a portion of the country programme budget for the conduct of a CPE. This was also the case in over 50% of COs in Africa, Asia and the Arab States regions. In the Eastern Europe and Central Asia region, neither of the COs had budgeted a CPE at the beginning of the programme cycle. The majority of COs were therefore required to comply with the request for a CPE without having made the necessary budget arrangements at an early stage of the programme.

The average cost of a CPE was $37,000, and ranged from $6,930 to $81,674. On average, this cost corresponded to 0.25% of the CP budget, ranging from a minimum of 0.04% (Peru) to a maximum of 0.8% (Tunisia and Panama). As a comparison, the 2005 UNFPA meta-evaluation reported that the cost of a good evaluation should represent 1 to 5% of the CP budget.[32]

---

[30] See 1.5.8 Meeting Needs.
[31] 2005 meta-evaluation, p.43.
[32] ibid, p.48.

2.1.2 Cost of other evaluations

Over half of COs participating in the CPE survey had conducted a mid-term review (MTR), for an average cost of $15,500. However, it appears from the 2012 EQA that the results from the MTR were often not used in the CPE process. This suggests that, despite their significant cost (almost 50% of the average budget of a CPE), the contribution of MTRs to subsequent evaluation work is limited.

COs reported that, during the time period of the CP cycle, over 110 other evaluations or reviews were conducted, which corresponds to an average of 2.75 per CO. This includes project reviews, UNDAF reviews, joint UN programme evaluations, and thematic evaluations. It is not possible to estimate the average cost of these additional evaluations and reviews as not all COs provided detailed budgetary information.

Diagram 6. Number of other evaluations or reviews conducted during CP cycle
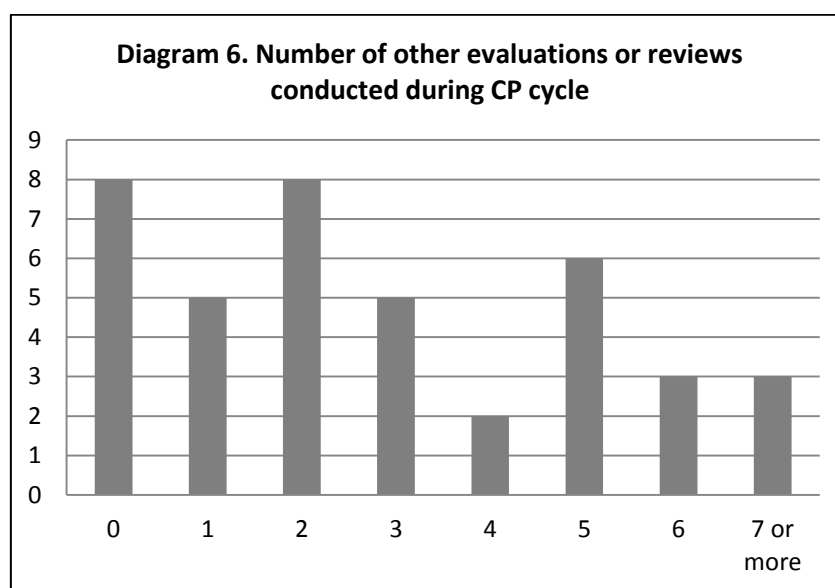
2.2 Management of Country Programme Evaluations

This section provides an analysis of responses to CPE survey questions on the manager of the evaluation, the M&E role in country offices, the drafting of the terms of reference for the evaluation, the formation and composition of a reference group, and the process for the approval of the final report and quality assurance.

2.2.1 The M&E role in country offices

There are two possible M&E roles in the CO: an M&E officer (a staff member working full-time on M&E duties), or an M&E focal point (a staff member who is also assigned other duties, such as national programme officer, or is a senior staff member such as the Assistant Representative).

Over two-thirds of COs had only an M&E focal point, which meant that the majority of COs lacked a staff member working full-time on M&E tasks. As an M&E focal point has a number of other professional responsibilities, it is difficult for these COs to dedicate enough time to ensure the quality of evaluations.

Diagram 7. The M&E role in country offices
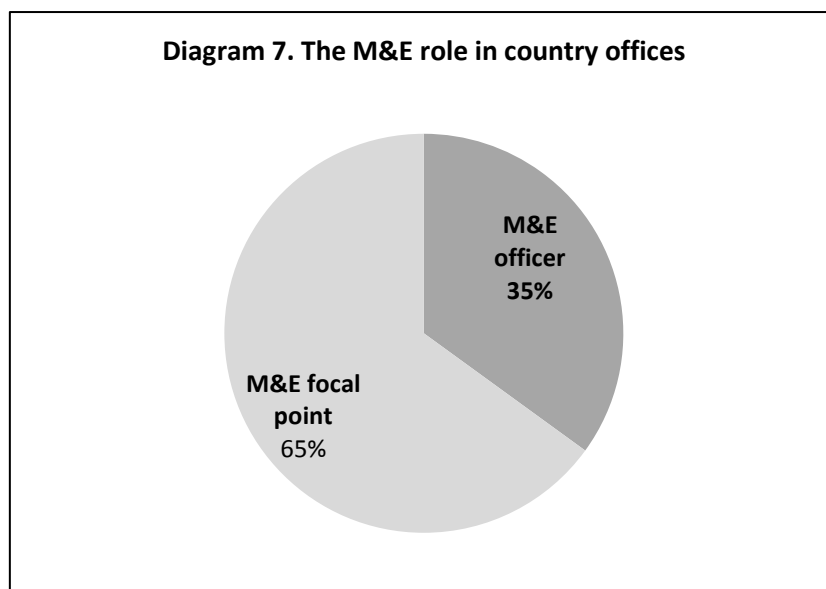
At a corporate level, UNFPA seems to be attaching growing importance to the level of quality of the COs M&E systems. UNFPA corporate policies explicitly recognise the core importance of results-oriented monitoring,[33] as "monitoring, evaluation and reporting constitute the foundation of UNFPA accountability to the Member States and the people they represent"[34] and "provides information that strengthens organisational decision-making and promotes a culture of accountability among programme implementers".[35] This emphasis stems from the understanding that the good functioning of the M&E system constitutes an essential basis for the CO to conduct adequate management and regular follow-up of the CP, including achieving a good CPE. The two independent CPEs recently conducted by the Evaluation Branch in Bolivia and Cameroon[36] both confirmed the negative consequence of the absence of a full-time M&E officer on the performance of the CO monitoring system.

Diagram 8 shows the CPE-related tasks assigned to the M&E officer/focal point. These cover all stages of the design, conduct and reporting of a CPE. As already stressed, in the case of the M&E focal point, these CPE tasks came in addition to their primary responsibilities. The management of the CPE was assured, in the majority of cases, by the M&E officer or focal point (63% of COs).[37]

---

[33] UNFPA Strategic Plan 2008-2011, paragraph 87-88 p.21; UNFPA Programme Manager's Planning, Monitoring and Evaluation Toolkit, paragraph i, vi, vii.

[34] UNFPA Strategic Plan 2008-2011, para 118 p.29.

[35] UNFPA Programme Manager's Planning, Monitoring and Evaluation Toolkit, paragraph vii.

[36] http://www.unfpa.org/public/home/about/oversight

[37] The UNFPA Representative or Deputy Representative also managed the evaluation for some CPEs.

Diagram 8. M&E officer/focal point involvement in CPE

- Drafting TOR
- Support to evaluation process
- Management of evaluation
- Quality control
- Editing final report
- Arranging meetings/providing documents

(28%, 28%, 21%, 20%, 2%, 1%)
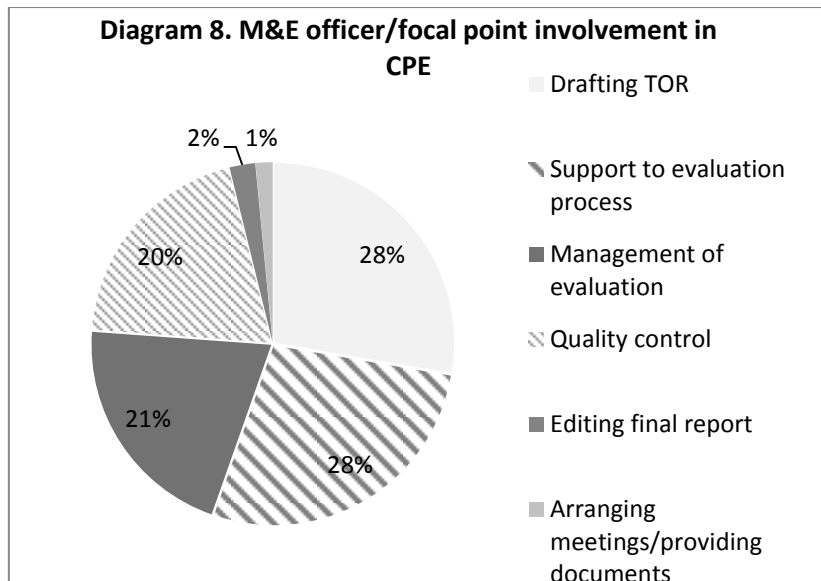
2.2.2 Drafting of the Terms of Reference for the CPE

The ToRs for the CPE were often drafted by the M&E officer/M&E focal point. Support was sometimes provided by another senior staff member from the CO, such as the UNFPA Representative, or Assistant or Deputy Representative.

Diagram 9. Personnel involved in drafting the Terms of Reference

- M&E Officer/Focal point
- M&E Officer/Focal point plus Asst/Dep Rep
- M&E Officer/Focal point plus International Programme Officer
- M&E Officer/Focal point plus Rep

(43%, 28%, 7%, 22%)
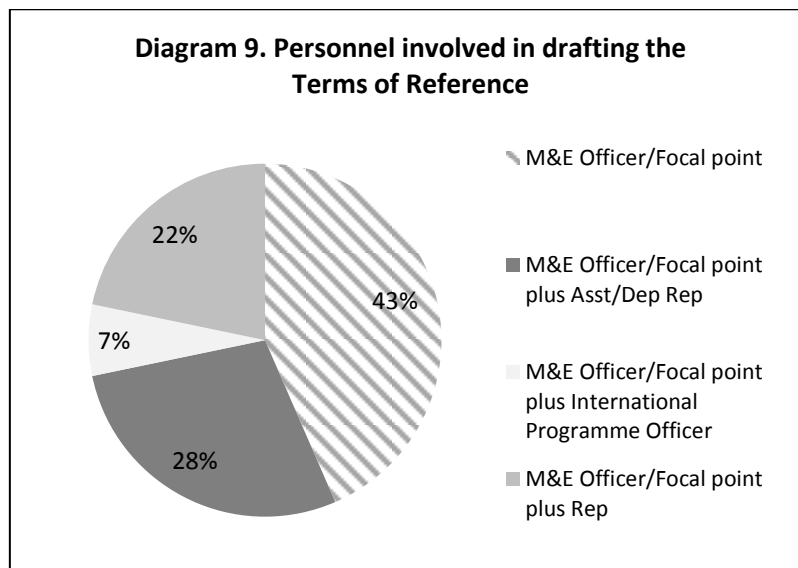
As indicated in the diagram below, regional office (RO) M&E advisors often provided assistance for the drafting of the ToRs. For almost 50% of the COs, the drafting of the ToRs involved government representatives. A more limited proportion of COs also invited other national and international counterparts to take part in the drafting of the ToRs and to become members of the CPE Reference Group.[38]

---

[38] See Section 2.2.3 Role of the Reference Group.

Diagram 10. Assistance provided for preparation of ToRs

A bar chart with a vertical axis ranging from 0 to 35 (in increments of 5) and the following categories on the horizontal axis: NGOs (approx. 5), Government (approx. 19), Academia (approx. 1), Agencies (approx. 8), Other (approx. 7), RO M&E (approx. 34).
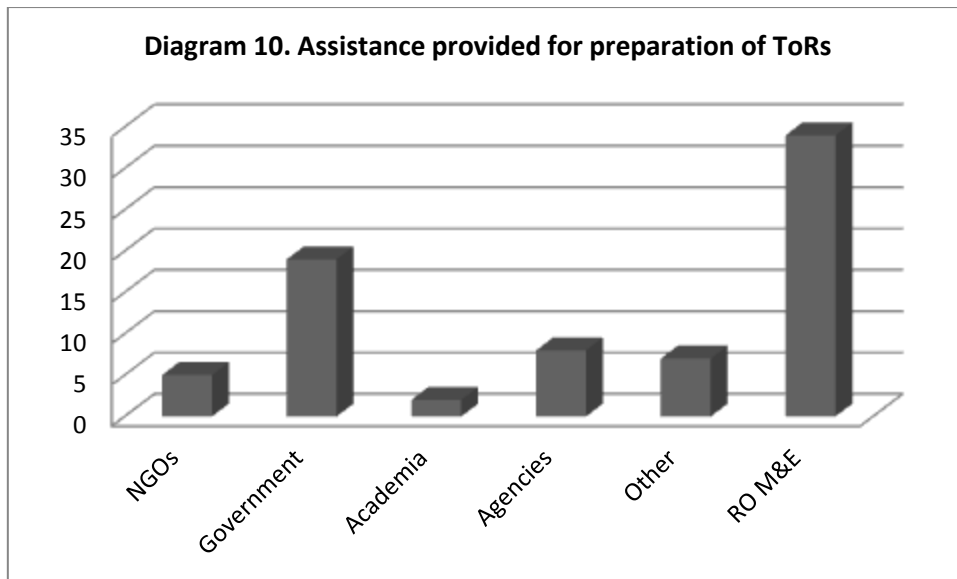
This involvement of multiple UNFPA staff members and external counterparts in the development of the ToRs proved useful, in particular for the COs that do not have a staff member working full-time on M&E tasks. Previous UNFPA evaluation quality assessments have noted that weaknesses in the quality of the ToRs can severely undermine the overall quality of the CPE.[39] Drawing on the expertise of other partners to draft the ToRs, as well as involving them in other key tasks such as participating in the reference group can help to raise the quality of the CPE. The participation of national stakeholders also contributes to developing national evaluation capacities.

2.2.3 Role of the Reference Group

A reference group, or steering committee or similar group, was established for almost all the CPEs.[40] However, the composition of reference groups varied from one CO to another. One CO reported that the reference group was "a team comprising one representative from government coordination ministry and the UNFPA M&E officer" whereas, in other instances, reference groups were larger groups incorporating national counterparts from a variety of organisations and several UNFPA programme staff. Two-thirds of CPE reference groups comprised government representatives.

---

[39] 2005 meta-evaluation, p.26.
[40] A number of COs reported that it was not possible to establish a formal Reference Group, for example as "there was not established a strong relationship with the new government, after the political crisis. However, consultations were made during the evaluation with the national associates". These are indicated in the above data chart as 'Ad-Hoc RG'.

Diagram 11. Reference Group composition (by member and frequency of their representation)
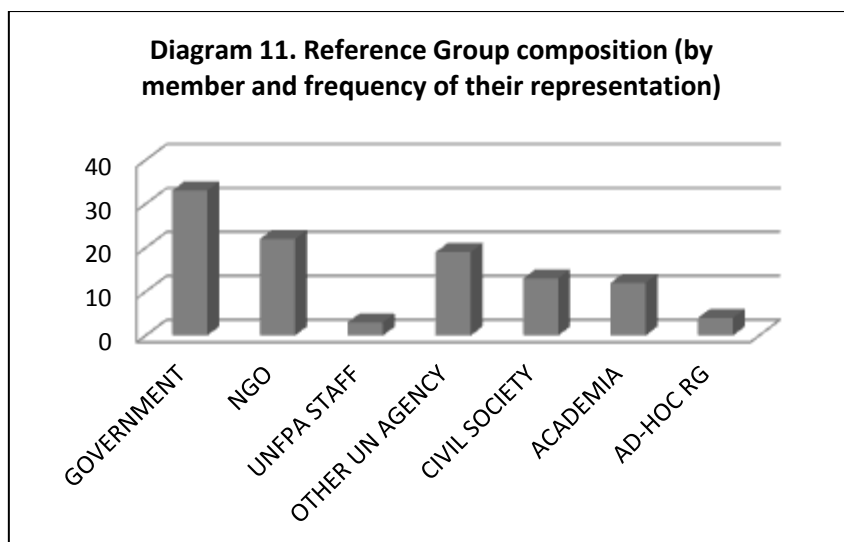
The tasks devolved to reference groups included, in descending order of frequency: drafting the ToRs; reviewing the final report; reviewing the composition of the evaluation team; reviewing the desk report; supervising the data collection. Almost all reference groups were involved in reviewing the final report and over two-thirds took part in preparatory tasks such as drafting the ToRs and reviewing the composition of the evaluation team.

2.2.4 Final CPE report and quality assurance

The assessment of the quality of the final report, prior to its approval, generally involved the reference group as well as regional offices M&E advisors.

The majority of COs (65%) stated that both the CO and the RO had conducted a quality assurance process, while the remainder (35%) indicated that the quality was checked by the CO alone . The results of the CPE survey did not indicate what resources were used for this process. Two-thirds of COs reported that they utilised a 'standardised' quality assessment template but provided no further detail.

2.3 Country Programme Evaluation Teams

This section touches upon the following issues: composition of the evaluation team, involvement of national and international consultants and the duration of the consultants' contracts.

Diagram 12. Composition of the core evaluation team: consultants and country office staff
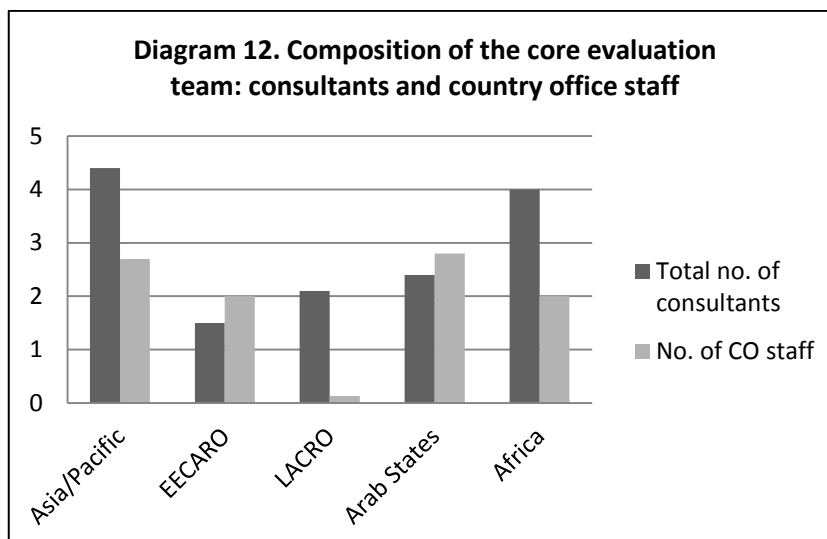
The level of involvement of the CO in the CPE process was reasonably consistent across all regions, with an average of two CO staff members.

Diagram 13. Composition of the core evaluation team: consultants
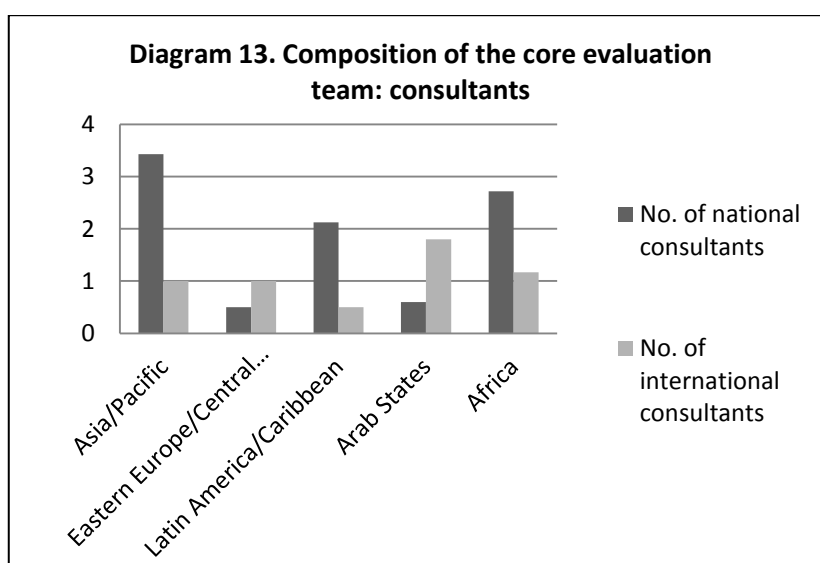
The use of consultants (and, in particular international consultants, given their higher cost) varied considerably from one CO to another, depending on the financial resources allocated to the CPE. This is particularly relevant given that the majority of CPEs were unbudgeted. COs in the Latin America and the Caribbean Region, which featured the highest proportion of unbudgeted CPEs,[41] indeed reported the lowest use of international consultants (see Diagram 13).

---

[41] See Section 2.1.2 Financing of the CPE.

Diagram 14. Availability of national consultants by type
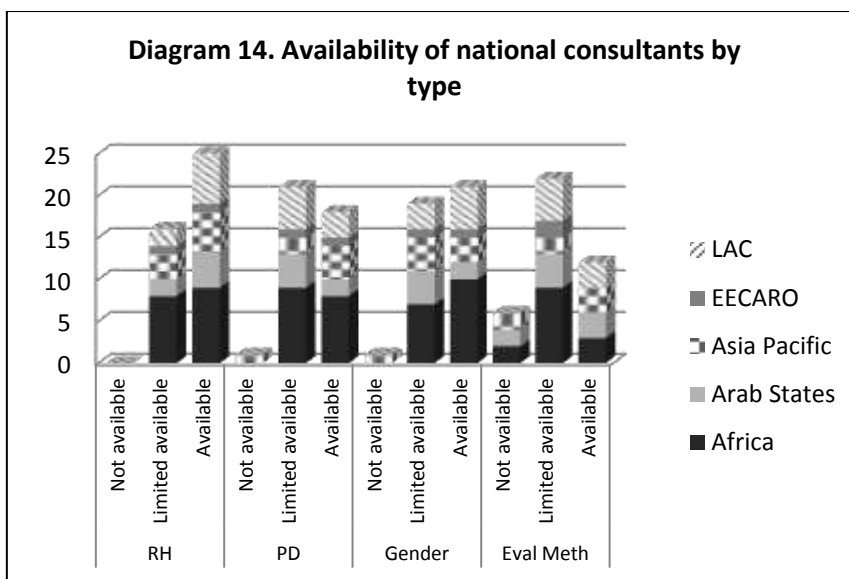
The availability of national consultants with relevant expertise varied between regions, which affected the composition of the evaluation team. National consultants with expertise in Reproductive Health or Gender were more available than other types of consultants in all regions. National consultants with experience in evaluation methodology were reported as being the most difficult to obtain across all regions.

Diagram 15. Duration of contracts
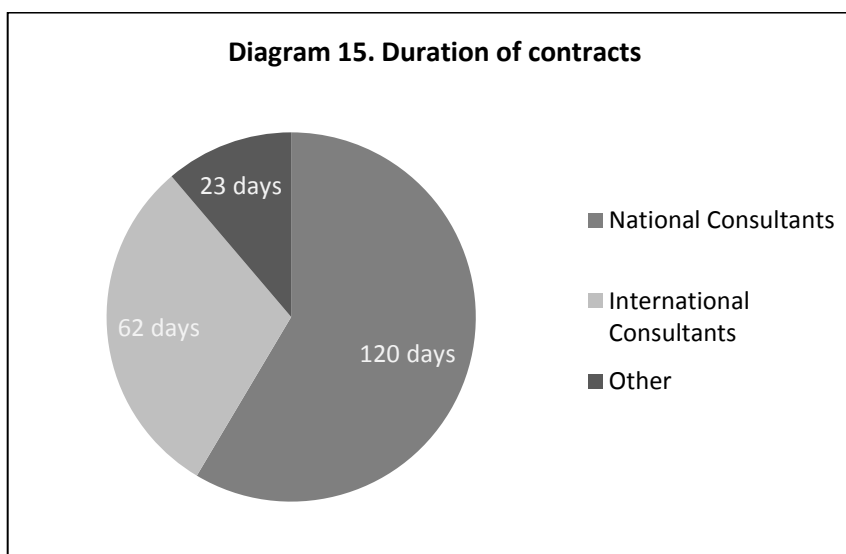
NB the 'Other' category corresponds to Research Assistants

The duration of the contracts of the consultants involved in CPEs present a diverse picture:
- UNFPA Iran reported contracts of a total of 330 days for national consultants and did not use any international consultants;
- UNFPA Guinea-Bissau hired 3 international consultants for 30 days each, and 3 national consultants working 45 days each;
- UNFPA Lao used no national consultants and used 3 international consultants with contracts of 42 days each.

Such differences between regions indicate the need for clearer guidance for COs on the optimal composition of the evaluation team. The 2005 meta-evaluation stated that "to achieve good quality evaluations, UNFPA country offices need to have access to capable national evaluators. Experience suggests that a mixed team consisting of national and international evaluators may prove to be the most optimal team".[42]

---

[42] 2005 meta-evaluation, p.49.

Part III. Overall Conclusions and Recommendations

3.1 Conclusions

Conclusion 1: Evaluation quality at UNFPA remains unsatisfactory

The 2012 EQA largely confirms previous quality assessments and shows that evaluation quality at UNFPA remains a pressing concern. Specifically, CPE reports lack sound and rigorous analytical work, thus resulting in insufficiently substantiated findings. Reports also suffer from the poor formulation of conclusions and the limited usefulness (due to lack of operational feasibility) of recommendations. As a consequence, results of the many CPEs now produced in COs cannot be used in confidence to inform the next programming cycle.

Conclusion 2: An increased number of CPEs without a parallel increase in COs resources (budget and staff)

Following the decision of the Executive Board that UNFPA programmes be evaluated at least once during their cycle, the sudden rise of CPEs has led to layers of evaluation and mid-term reviews stacked together, yet not appropriately planned, so that MTRs are of little (if any) use for the conduct of CPEs. Furthermore the addition of MTRs and CPEs has resulted in a strain on already limited financial and human resources that COs can possibly dedicate to evaluation exercises. As a result, since they come towards the end of the cycle, the CPEs were often allocated budgets and staff that were largely insufficient in view of achieving a good quality report.

Conclusion 3: Absence of efforts with regard to the development of evaluation capacities in COs further limits the quality of CPEs

COs have formally complied with the Executive Board request to have all country programmes evaluated at least once in a cycle. However the staff assigned to the management of CPEs in COs often lack the required expertise and experience. Little has been done with regard to the development of evaluation capacities in COs. This negatively affects the quality and credibility of CPE reports – hence significantly reducing their usefulness for informing the next programming cycle.

Conclusion 4: The absence of results-oriented monitoring impairs the quality of CPEs

In the absence of an effective results-oriented monitoring system, country programme action plans and associated results frameworks suffer from considerable deficiencies which, in turn, impair the quality - and therefore the credibility and usefulness - of CPEs. Monitoring is indeed an indispensable complement to all evaluation activities as it should provide: (i) information on the programme implementation progress and (ii) the set of indicators against which evaluators must assess the programme performance.

3.2 Recommendations

Recommendation 1: Rationalize the conduct of mid-term review and evaluation exercises

Conducting both the country programme mid-term review and final evaluation within one single cycle has proved unmanageable for COs. This calls for a rationalization of these exercises with a view to:

    (i)   Ensuring more resources for the completion of the CPE;

    (ii)   Optimally articulating MTRs and CPEs, thus turning currently competing assignments into complementary exercises.

This rationalization could be performed through the undertaking of a CPE every second cycle while an MTR would be conducted for the programme cycle for which a CPE is not scheduled.

Recommendation 2: Ensure that COs are equipped with the necessary methodological skills

Ensure the wide dissemination and effective use of the *Handbook on How to Design and Conduct a Country Programme Evaluation at UNFPA.*

Recommendation 3: Set up a results-oriented monitoring both at corporate and country levels

Headquarters and the Regional Offices should:

- Promote the development and use of corporate guides and tools with of view to developing results-oriented monitoring capacities.
- Guarantee that funds are dedicated to the setting up of results-oriented monitoring systems in country offices.

Bibliography

Division for Oversight Services, *Evaluation Quality Assessment: 2007-2008*, United Nations Population Fund, New York, 2009.

Evaluation Branch, Division for Oversight Services, *Handbook on How to Design and Conduct a Country Programme Evaluation*, United Nations Population Fund, New York, 2011.

Executive Board of the United Nations Development Programme and United Nations Population Fund, *UNFPA Evaluation Policy*, United Nations Population Fund, New York, 2009.

Executive Board of the United Nations Development Programme and United Nations Population Fund, *Compendium of decisions adopted by the Executive Board at its annual session 2009 (26 May to 3 June 2009),* United Nations Population Fund, New York, 2009.

United Nations Population Fund, *The Programme Managers Planning, Monitoring and Evaluation Toolkit*, United Nations Population Fund, New York, 2004.

United Nations Population Fund, *Strengthening Evaluation for Improved Programming: UNFPA Evaluation Quality Assessment*, UNFPA, New York, 2005.

United Nations Population Fund, *UNFPA Strategic Plan 2008-2011: Accelerating Progress and National Ownership of the ICPD Programme of Action*, United Nations Population Fund, New York, 2007.

Annex 1. EQA grid explanatory notes and template

Origin and Purpose of the EQA

The introduction, as of May 2011, of the Evaluation Quality Assessment grid (EQA) was driven by the willingness to come on par with similar best practices enforced by our sister agencies and other international organizations and bilateral donors.

The main purpose of the EQA is to ensure that evaluation reports comply with professional standards while meeting the information needs of their intended users.

Use of the tool

The EQA grid should be filled twice by the evaluation manager, at the last two stages of the evaluation process.

The first time corresponds to the assessment of the draft final evaluation report. The grid is mainly directed to the evaluation team, with a view to pointing out areas of improvement for the final version of the report.

The second assessment concerns the final report and consists in expressing a final quality judgment on the evaluation report. This final EQA grid is the one that will be sent to Headquarters (HQ) management, along with the evaluation final report, to be published in the database of the Evaluation Branch at DOS. The audience targeted is HQ management but also the general public, as the grid is made available on the Internet. Internally (at UNFPA), the assessment of the strengths and weaknesses of the evaluation report gives an indication of the relative reliability of its results and determines the extent to which the report (or elements of it) can be used to feed lessons learned into future programming.

From an external perspective, publishing the final EQA grid together with the evaluation report contributes to the transparency and credibility of UNFPA when reporting on its performance.

Structure of the document

The first part of the document (Part I) presents explanations on the content of the EQA grid as well as indications on how to fill it. The second part (Part II) corresponds to the template to be used by the evaluation managers.

1. Explanations regarding the Quality Assessment criteria

<div style="border:1px solid">

1. Structure and Clarity of Reporting

Does the report clearly describe the evaluation, how it was conducted, the findings of the evaluation, and their analysis and subsequent recommendations?

Is the structure *logical*? Is the report *comprehensive*?

Can the information provided be *easily understood*?

</div>

2. Executive Summary

Does it read as a stand-alone section, and is a *useful* resource in its own right?

Is it brief yet *sufficiently detailed*, presenting the main results of the evaluation, and including *key elements* such as methodology and conclusions and recommendations?

3. Design and Methodology

Is the *methodology* used for the evaluation clearly described and is the rationale for the methodological choice justified?

Have cross-cutting issues (vulnerable groups, youth and gender equality) been paid specific attention (when relevant) in the design of the evaluation?

Are key processes (tools used, triangulation, consultation with stakeholders) discussed in sufficient detail? Are *constraints* and *limitations* made explicit (including limitations applying to interpretations and extrapolations; robustness of data sources, etc.) and discussed?

4. Reliability of Data

Are *sources* of data clearly stated for both primary and secondary data?

Is it clear why case studies were selected and what purpose they serve?

Are all relevant materials related to case studies, interviews (list of interviewees, questionnaires) etc. annexed to the report?

Are the limitations, and methods to address them, discussed?

What other *data gaps* are there and how have these been addressed?

5. Findings and Analysis

Findings
Is there a *clear pathway* from data to findings, so that all findings are *evidence-based*?

Are *biases* stated and discussed?

Are *unintended* findings reported and discussed?

<u>Analysis</u>
Are *interpretations* of the findings understandable? Are *assumptions* clearly stated and extrapolations well explained?

Are their *limitations* (or drawbacks) discussed?

Does the analysis respond to <u>all</u> evaluation questions?

If not, are *omissions* (of both evaluation criteria and questions) recognized and explained?

Has the analysis examined *cause and effect* links between an intervention and its end results?

Are *contextual factors* identified and their influence discussed?

6. Conclusions
Are the conclusions organized in priority order?

Do the conclusions amount to a reasonable *judgment* of the findings and are their links to evidence made clear?

Are there any limitations and are these made clear?

Do they present an *unbiased* judgment by the evaluators of the intervention or have they been influenced by preconceptions or assumptions that have not been discussed?

7. Recommendations
Is there a *logical flow* from the conclusions to recommendations?

Are they strategic and clearly presented in a priority order which is consistent with the *prioritization* of conclusions? Are they *useful* – sufficiently detailed, targeted and likely to be implemented and lead to *further action*?

How have the recommendations *incorporated* stakeholders' views and has this affected their *impartiality*?

8. Meeting Needs
Does the report adequately address the information needs and responds to the *requirements stated in the ToRs?*

In particular does the report respond to the evaluation questions, issues or criteria identified in ToR?

[Insert organizational name]: EQA for [insert title of assessed report]

This version of the Evaluation Quality Assessment (EQA) grid template (and related Explanatory Note) must be attached to the ToRs of all evaluations commissioned by UNFPA. Upon receipt of the final Evaluation Report, the Evaluation Manager must complete an EQA using this template and upload it, along with the final Evaluation Report in Docushare.

Title of Evaluation Report:
Name of Evaluation Manager:
Name of EQA Reviewer (if different to above):
Budget and time frame allocated for this evaluation:

Overall Assessment: Note that the overall assessment must address, as a minimum, the following issues: *scope of the evaluation; methodological design; findings and analysis; credibility of data; recommendations; conclusion; executive summary.*

Quality Assessment criteria	Assessment Levels			
	Very Good	**Good**	**Poor**	**Unsatisfactory**
1. Structure and Clarity of Reporting *To ensure report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.* Checklist of minimum content and sequence required for structure: • i) Acronyms; ii) Exec Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned (where applicable) • Minimum requirements for Annexes: ToRs; Bibliography List of interviewees; Methodological instruments used.	*Please insert <u>assessment level</u> followed by your main <u>comments</u>.*			

2. Executive Summary *To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation.* Structure (paragraph equates to half page max): • i) Purpose, including intended audience(s); ii) Objectives and Brief description of intervention (1 para); iii) Methodology (1 para); iv) Main Conclusions (1 para); v) Recommendations (1 para). Maximum length 3-4 page	
3. Design and Methodology *To provide a clear explanation of the following elements/tools* Minimum content and sequence: • Explanation of methodological choice, including constraints and limitations; • Techniques and Tools for data collection provided in a detailed manner; • Triangulation systematically applied throughout the evaluation; • Details of participatory stakeholders' consultation process are provided. • Whenever relevant, specific attention to cross-cutting issues (vulnerable groups, youth, gender equality) in the design of the evaluation	
4. Reliability of Data *To clarify data collection processes and data quality* • Sources of qualitative and quantitative data have been identified; • Credibility of primary (e.g. interviews and focus groups) and secondary (e.g. reports) data established and limitations made explicit.	
5. Findings and Analysis *To ensure sound analysis and credible findings* <u>Findings</u> • Findings stem from rigorous data analysis; • Findings are substantiated by evidence; • Findings are presented in a clear manner <u>Analysis</u> • Interpretations are based on carefully described assumptions; • Contextual factors are identified. • Cause and effect links between an intervention and its end results (including unintended results) are explained.	

6. Conclusions *To assess the validity of conclusions* • Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention.	
7. Recommendations *To assess the usefulness and clarity of recommendations* • Recommendations flow logically from conclusions; • Recommendations must be strategic, targeted and operationally-feasible; • Recommendations must take into account stakeholders' consultations whilst remaining impartial; • Recommendations should be presented in priority order	
8. Meeting Needs *To ensure that Evaluation Report responds to requirements (scope & evaluation questions/issues/DAC criteria) stated in the ToR (ToR must be annexed to the report).* *In the event that the ToR do not conform with commonly agreed quality standards, assess if evaluators have highlighted the deficiencies with the ToR.*	

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Unsatisfactory	Poor	Good	Very good
5. Findings and analysis (50)				
6. Conclusions (12)				
7. Recommendations (12)				
8. Meeting needs (12)				
3. Design and methodology (5)				
4. Reliability of data (5)				
1. Structure and clarity of reporting (2)				
2. Executive summary (2)				
TOTAL				

(*) Insert the multiplying factor associated with the criteria in the corresponding column e.g. - if "Finding and Analysis" has been assessed as "good", please enter the number 50 into the "Good" column. The Assessment level scoring the higher number of points will determine the overall quality of the Report

OVERALL QUALITY OF REPORT: [Insert overall Assessment Level based on highest score above – see Explanatory Note for further guidance and example]

Annex 2. EQA for UNFPA country programme of [country X]

Title of Evaluation Report: Evaluation of the UNFPA Country Programme to Country X 2007-2010

Overall Assessment: The report's **structure** includes all main elements but is too lengthy in key sections and the structure does not aid the reader in gaining a clear overview of the evaluation's results. Very little information or detail on the **methodology** is provided. Multiple sources of **data** are utilized but referencing is weak and there is low use of a number of other relevant evaluations. The **findings** section is lengthy in the description of activities but it is often unclear if this refers to UNFPA. There is little use of evidence to justify statements made about achievements. As a consequence, the **conclusions** struggle to synthesis the findings but do partially achieve this and are able to communicate the evaluators' judgments. **Recommendations** are poorly-written and lack sufficient detail to be of use to the CO. Overall the report is not well-integrated and does not meet the needs laid out in the **ToRs**.

Quality Assessment criteria	Assessment Levels			
	Very Good	**Good**	**Poor**	**Unsatisfactory**
1. Structure and Clarity of Reporting *To ensure report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.* Checklist of minimum content and sequence required for structure: • i) Acronyms; ii) Exec Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned (where applicable) • Minimum requirements for Annexes: ToRs; Bibliography List of interviewees; Methodological instruments used.	**Poor** The report has a conventional structure with detailed annexes. However clearer information on context would have been useful as details on areas selected for projects, justifying the selection of these locations, is not provided until mid-way through the findings section.			
2. Executive Summary *To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation.* Structure (paragraph equates to half page max): Max length 3-4page	**Good** The Executive Summary provides clear background and context to the CPE. The section is brief, highlighting in short paragraphs key findings and conclusions although it would benefit structurally from			

• i) Purpose, incl. intended audience(s); ii) Objectives & Brief description of intervention (1 para); iii) Methodology (1 para); iv) Main Conclusions (1 para); v) Recommendations (1 para).	subheadings and from a clearer indication of the key recommendations. However the executive summary functions well as a standalone section.
3. Design and Methodology *To provide a clear explanation of the following elements/tools* Minimum content and sequence: • Explanation of methodological choice, including constraints and limitations; • Techniques and Tools for data collection provided in a detailed manner; • Triangulation systematically applied throughout the evaluation; • Details of participatory stakeholders' consultation process are provided. • Whenever relevant, specific attention to cross-cutting issues (vulnerable groups, youth, gender equality) in the design of the evaluation	**Poor** Little is said about the methodological choices made. In particular, the choice of the four regions selected for in-depth study could have been better explained (although it seems relevant). Difficulties and constraints (in particular as regards the scarcity and absence of baseline data) are made clear but lack comment on ways to overcome them. The evaluation relies almost entirely upon documentary analysis and interviews, whereas other tools are necessary. Case studies, in particular, would have been highly useful in a context of scarcity of data. Limitations of the tools used are not mentioned. There is also no evidence of triangulation.
4. Reliability of Data *To clarify data collection processes and data quality* • Sources of qualitative and quantitative data have been identified; • Credibility of primary (e.g. interviews and focus groups) and secondary (e.g. reports) data established and limitations made explicit;	**Poor** The report contains little data and in particular lacks data regarding the outputs, results and impacts of the programme. Although the evaluators rightly point out the absence of baseline data, there is no evidence of their attempt to establish or reconstruct these data. Where data is provided, it mainly concerns resources and budgetary execution of the programme and sources are well identified and appear reliable (ATLAS financial report).
5. Findings and Analysis *To ensure sound analysis and credible findings* Findings • Findings stem from rigorous data analysis; • Findings are substantiated by evidence; • Findings are presented in a clear manner	**Poor** Findings are not evidence based, mainly consisting of statements which are insufficiently substantiated, and often reference is made to interviews as the only source of information. Again, triangulation is not evident from the analysis. Findings are presented in a mechanical way around the DAC criteria.

<u>Analysis</u> • Interpretations are based on carefully described assumptions; • Contextual factors are identified. • Cause and effect links between an intervention and its end results (including unintended results) are explained.	Furthermore, findings are mainly at the level of activities, thus effectiveness of results is also a question mark. There is no clear assessment of whether effects will be sustainable over time - this is particularly evident on the section of population and development. Contextual factors are not well described and the cause/effect links are not well explained.
6. Conclusions *To assess the validity of conclusions* • Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention.	**Poor** Conclusions and Recommendations are combined in a single chapter, by mandate area. The detailed findings have not been translated into effective conclusions, and it is difficult to distinguish what were the achievements or shortcomings of the CP that were outlined in the findings section.
7. Recommendations *To assess the usefulness and clarity of recommendations* • Recommendations flow logically from conclusions; • Recommendations must be strategic, targeted and operationally-feasible; • Recommendations must take into account stakeholders' consultations whilst remaining impartial; • Recommendations should be presented in priority order	**Poor** The report contains a clustered set of recommendations, covering well all the dimensions of the programme. They address the programming stage, the implementation, monitoring and evaluation and also identify possible areas of intervention to be considered in the next programming cycle. However, these recommendations are not related to conclusions (which do not exist as such) and can only partially be deduced from findings. They also lack prioritization and options regarding their implementation.
8. Meeting Needs *To ensure that Evaluation Report responds to requirements (scope & evaluation questions/issues/DAC criteria) stated in the ToR (ToR must be annexed to the report).* *In the event that the ToR do not conform with commonly agreed quality standards, assess if evaluators have highlighted the deficiencies with the ToR.*	**Poor** The ToRs include a lot of evaluation questions, including Impact, and evaluators have not made any substantive comments on weaknesses in the ToRs. The final evaluation report has not addressed all criteria under the ToRs (or explained the reason why it can't be addressed). Therefore the report does not meet the needs laid out in the ToRs.

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Unsatisfactory	Poor	Good	Very good
5. Findings and analysis (50)		50		
6. Conclusions (12)		12		
7. Recommendations (12)		12		
8. Meeting needs (12)		12		
3. Design and methodology (5)		5		
4. Reliability of data (5)		5		
1. Structure and clarity of reporting (2)		2		
2. Executive summary (2)			2	
TOTAL		98	2	

(*) Insert the multiplying factor associated with the criteria in the corresponding column e.g. - if "Finding and Analysis" has been assessed as "good", please enter the number 50 into the "Good" column. The Assessment level scoring the higher number of points will determine the overall quality of the Report

OVERALL QUALITY OF REPORT: Poor

Annex 3. 2012 EQA: results by individual assessment criteria

Assessment Level	Overall	Structure and Clarity of Reporting	Executive Summary	Design and Methodology	Reliability of Data	Findings and Analysis	Conclusions	Recommendations	Meeting Needs
Unsatisfactory	8	2	7	7	5	8	15	5	6
Poor	23	22	18	19	21	23	16	23	19
Good	3	10	8	7	8	3	3	6	9
Very Good	0	0	1	1	0	0	0	0	0

Annex 4. 2012 EQA: Overall Assessment Summary for EQAs performed by the Evaluation Branch

The UNFPA Evaluation Quality Assessment (EQA) aims to establish the status of and improve evaluation report quality in UNFPA by reviewing and rating the quality of Country Programme Evaluations performed by, or on behalf of, UNFPA. It is managed by DOS Evaluation Branch to ensure independence of the quality process.

All CPE reports eligible for review are assessed by DOS Evaluation Branch on 8 criteria and assigned a rating of 1 of 4 assessment levels (Unsatisfactory, Poor, Good, Very Good). These ratings are then combined to provide an overall assessment level for the report.

The countries whose CPE was assessed by Evaluation Branch in 2011 are:
Algeria, Bangladesh, Brazil, Central African Republic, Chad, Dominican Republic, Egypt, El Salvador, English/Dutch-speaking Caribbean, Eritrea, Gabon, Gambia, Ghana, Honduras, Iran, Kyrgyzstan, Lao PDR, Malawi, Mauritania, Mongolia, Mozambique, Morocco, Myanmar, Panama, Peru, Philippines, Sao Tome & Principe, Senegal, Syrian Arab Republic, Tanzania, Thailand, Ukraine, Republic of Yemen and Zimbabwe.

All Evaluation reports are located in the UNFPA Evaluation Database at: http://web2.unfpa.org/public/about/oversight/evaluations/

Country Name	Assessment (Summary)
Algeria	The present document does not qualify as a CP evaluation report. Instead of structuring the report along a set of evaluation questions and/or DAC criteria (as suggested in ToR), the scope of the assessment consists of a broad screening of activities under the 3 mandate areas. As such, this report amounts to a type of "implementation monitoring" exercise which does not result in any added value. The report does not provide any indication that a **methodological design** was developed and it seems that only a documentary review and a few interviews (the annex lists only 12 interviewees) were implemented. As a result, **data** is scarce and the report's **findings and analysis** are entirely confined to a rapid assessment of activities without any discussion of progress towards results. **Conclusions** and **recommendations** are respectively largely limited to a few human resources and management concerns or long developments on a couple of possible priority areas for the next CP. The very unsatisfactory quality of the report (both form – e.g.; it does not contain an **Executive Summary**, and substance) is particularly striking in view of the **ToRs** whose demands in terms of questions (and related DAC criteria), focus (on results) and methodology, although not very detailed, were sufficiently clear.

Bangladesh	The report's **structure** is comprehensive and the **Executive Summary** provides an overview of the evaluation's **conclusions** but lacks detail. **Methodology** and tools are well-described, and the evaluators' decision to focus on activities and processes rather than results is explained as a response to concerns about limitations of the Results and Resource Framework, and the decision not to utilize the revised RRF is also explained. However the evaluation is limited by the lack of results and the decision not to recreate missing baselines is not adequately explained. The **findings** are presented in a lengthy format that does not read in a consistent manner or enable the reader to reach **conclusions**. The **ToRs** contain extensive detail on requirements and the inclusion of impact evaluation is beyond the means of the proposed methodology. The **recommendations** do not present solutions to identified problems but instead solely focus on the 8th CP. The evaluation meets the needs of the ToRs which were broad yet detailed but included elements which the evaluation team would have benefitted from excluding.
Brazil	The report uses plain language but fails to communicate effectively by being too repetitive. Its **structure** does not correspond to usual standards. It satisfactorily covers the evaluation scope indicated in the ToRs**.** ToRs did provide guidance to the evaluators; they do make reference to the DAC criteria although they wrongly refer to the criteria as objectives. The main **Executive Summary** is clear and satisfactorily concise. The **methodological design** is insufficiently presented and explained with no evidence of an actual triangularisation of **data** sources; limitations have been well identified though. Since sources are not referred to, it is very difficult to assess the reliability of the data used, plus information on baselines is not provided. The absence of substantiating evidence explains the limited credibility of **findings**. As regards **recommendations**, they are balanced and presented in clusters and are sufficiently operationalized to be implemented. However they do not refer to corresponding conclusions because there are no **conclusions** as such. Overall, the report lacks a clear and explicit logical flow from findings to conclusions to recommendations nor does it respond to the requirements stated in the **ToRs** in particular to the questions on effectiveness and sustainability of the effects of the Programme.
Central African Republic	The report satisfactorily covers the evaluation scope indicated in the ToRs. However, the **ToRs** should have provided the evaluators with more guiding information such as a clear reference to the DAC criteria and the necessity to structure the evaluation around a clear and limited set of evaluation questions. Evaluation results are well synthesised in a concise and clear **Executive Summary**. The **methodological design** (and its limitations) is insufficiently presented and explained. In particular, there is no evidence of an actual triangularisation of data sources. Moreover, since these sources are scarcely (if ever) referred to, it is very difficult to assess the reliability of the **data** used. The absence of a rigorous analysis and of substantiating evidence account for the (very) limited credibility of **findings**. Finally, the report lacks a clear and explicit logical flow successively leading from findings to **conclusions** and to recommendations. As regards **recommendations**, they are not presented in prioritized clusters and their operationalization is limited.

Chad	The lengthy **Executive Summary** serves its purpose of providing an overview of the evaluation's scope and objectives, methodological choices/constraints and main results. However, the scope of the evaluation is not sufficiently focused particularly in view of the fact that the exercise was to be performed in 6 weeks only (as per **ToRs**). Instead of structuring the evaluation along a set of limited evaluation questions, the report presents a screening of all 3 mandate areas by DAC criteria (including impact which should have been excluded given the fact that the evaluation takes place a year before the end of the CP). The **methodological design** is unsatisfactory with only: (i) a few tools the choice of which is, furthermore, not explained and (ii) a limited number of documents reviewed and interviewees (among whom, it is not clear that evaluators actually met with beneficiaries). Besides there is: no indication that sources were actually triangulated; no evaluation questions matrix; and limitations (although obvious) are not discussed. The report's **findings and analysis** section is entirely confined to an assessment of activities and does not present any discussion of progress towards CPAP results. In fact, the report amounts more to some sort of monitoring rather than a fully-fledged CP evaluation which probably was not feasible in such a short period of time. Although informative, these "monitoring findings" come too late into the CP implementation and have not led to the production of substantive **conclusions**. The **recommendations** are too many, not prioritized and insufficiently specific or actionable.
Dominican Republic	The report is not user-friendly, comprehensive or well-**structured**. There is no context section and the annex only contains a list of interviewees but contains no **ToRs**, no bibliography and no methodological tools. The **Executive Summary** is weak and incomplete with many sections missing. A description of the methodology utilized is also missing and there is no mention of beneficiaries. **Findings**, conclusions, lessons learned and recommendations are all mixed up adding to lack of clarity of the Executive Summary. **Methodological** section is weak and lacks clarity. Methodological choice, including constraints and limitations are not clearly explained or are missing. **Data** sources and data collection methods are not clearly explained. Triangulation is not mentioned or applied systematically throughout the evaluation. Very few stakeholders have been consulted and there is no stakeholder map. Issues of data quality and credibility are not discussed. There is no mention of triangulation or other strategies to improve reliability and validity of data. The issue of adequate baselines, indicators of progress and targets is not appropriately addressed. There is no discussion of how progress towards results is measured. There is a section at the end of the report called epilogue that contains **conclusions**. These are not organized in a priority order and some **recommendations** are mixed in with these conclusions. Some recommendations are clear, practical and strategic although they are in some cases based on findings and not conclusions. Some recommendations are mixed with findings and are therefore difficult to understand. Recommendations are not presented in priority order. Report does not mention **ToRs** and they are not part of the report.
El Salvador	The scope of the evaluation is not sufficiently focused especially in view of the fact that: (i) this evaluation does not focus on evaluation questions/issues by focus areas but rather on evaluation criteria; furthermore not all five criteria are evaluation criteria: "coordination" and "monitoring"; (ii) the evaluation does not address a very important evaluation criteria which is sustainability or potential sustainability of effects/benefits over time. The **methodological design**, tools

	are systematically presented in the text and examples of tools are included in the annexes yet triangulation approaches are not clearly explained. Moreover, the link between the evaluation criteria and the evaluation questions presented in the **ToRs** is not evident. The design is very much oriented around the so called "evaluation criteria" without much focus on the main outcomes by focus area which undermines the understanding of the extent to which results have been achieved, partially achieved or not achieved; Constraints/limitations are not mentioned in the Methodology section. Evaluators' **findings** appear to be based on **credible data** however the number of interviews is very slim and hardly any beneficiaries have been consulted. **Conclusions** are presented on a very simplistic way without a clear link to findings; judgment presented in the conclusions is not based in evidence. **Recommendations** do not flow from conclusions; they are not ranked in order of priority nor targeted. Yet they present some valid inputs that may be taken into account for the next programming cycle.
Egypt	The report's **structure** is not systematic, and is extremely poorly presented and difficult for the reader, and the **executive summary** fails to provide the reader with a clear overview of the evaluation results. The **methodology** is scattered across several sections as are **findings**, and significant quantities of relevant info are presented in annexes but not referred to in the main body of the report. The evaluators have attempted to identify sources of, and collect, **data** but its analysis is poor, with little evidence utilized for statements of results. The lack of clear, summarizing **conclusions** and a misuse of lessons learned leaves the reader with no clear summary, and **recommendations** are too numerous and non-strategic to provide a clear perspective on the CP. The report fails to meet the needs specified in the **ToRs** by neglecting to answer the evaluation questions.
English & Dutch-Speaking Caribbean Countries	The report's **structure** is not balanced, with a long **Executive Summary** (10 pages). The **methodology** fails to sufficiently address the severe weakness of evidence arising from the lack of baseline & end line **data**. The scope of the evaluation (multiple COs and partners) is large and presents a data collection challenge; however stakeholder engagement was a mixed success, with some obvious gaps and inappropriate use of partners' views to measure CP success. There is a lack of use of evidence for the **findings** and analytical approach is not consistent, with inconsistent use of evidence. **Conclusions** are not prioritized. However, practical **recommendations** have been produced. The lack of data throughout the evaluation has severely compromised the evaluators' ability to meet the **ToRs**.
Eritrea	The scope of the evaluation is insufficiently focused because of the absence of a clear set of evaluation questions. Very little (if anything) is said about the **methodological design** of the evaluation; apart from a generic list of evaluation tools, no explanation can be found regarding the methods and tools chosen, the methodological constraints and limitations faced and the way to overcome them. Above all, the report fails to present the reader with satisfactory analysis, leading to credible **findings**. The information provided is mainly descriptive, and although the **data** used are sufficiently reliable, they concern issues of relevance for monitoring purpose only (description of financial data, indicators at the level of activities and, more scarcely, outputs). The report lacks an actual section on **conclusions**. The absence of linkage with corresponding conclusions and the lack of information regarding their operational feasibility severely limit the usefulness

	of **recommendations**. Furthermore, the report lacks clarity and a logical structure. Its **Executive Summary** is far too long, while the absence of important information prevents it from being a self-standing document.
Gabon	The scope of the evaluation is not sufficiently focused especially in view of the general deficit of data in Gabon and the fact that the exercise was to be performed in 2 months only (as per **ToR**). As a result, instead of structuring the evaluation along a set of limited evaluation questions, the report presents a screening of all 3 mandate areas by DAC criteria (to the – appropriate - exclusion of impact) and other issues (such as: management, evaluation, partnership, etc.). The **methodological design** is poor (few tools the choice of which is, furthermore, not explained; no indication of triangulation of sources; no evaluation questions matrix; limitations are not discussed) and it does not address the issue of scarcity of **data** - notably the lack of baselines. The report's **findings and analysis** section is entirely confined to an assessment of activities and presents neither discussion of progress towards results nor **conclusions**. The **recommendations** are too many and not prioritized. Instead they are presented in 5 lists (the 3 mandate areas; recommendations to the GoG; to UNFPA, and to IPs) of actions which are neither sufficiently specific nor actionable. The lengthy **Executive Summary** does serve its purpose of providing an overview of the evaluation's scope and objectives, methodological choices/constraints and main results.
Gambia	The **Executive Summary** provides a concise overview, but does not include the recommendations or a concluding section. The **methodological approach** is well explained and supporting documentation provided in the annex. However, findings are not based on evidence and analysis lacks credible **data**. The report outlines best practices and lessons learned, however, it omits an explicit **conclusions** section, with the reader limited to a best practices and Lessons Learned section. **Recommendations** are therefore not easily linkable to respective findings and sometimes appear as 'stand-alone' recommendations. The **ToRs** contain many evaluation questions and present a challenge to the evaluators who have not responded with the required depth to the evaluation objectives.
Ghana	The report's unconventional **structure** is based around 'Goal and Outcome Level Assessment' and 'Output Level Assessment', and hampers readability. The **Executive Summary** is overly long and unfocused. **Methodology** is poorly described, with little information provided about **data** sources or collection methods, with no tools made available in the report. **Findings** are detailed, but often unreferenced, and where references are provided they are too reliant on a limited number of national reports, with no reference to what original data collection has taken place. The lack of use of indicators hampers analysis and the evaluators have utilised their own proxies, such as IP implementation rate, which fails to provide a clear picture of output results. There is no clear **conclusions** section. **Recommendations** are poorly presented, not prioritized and impractical. No **ToRs** are provided so it is not possible to determine whether the weaknesses in the structure and content of the report are due to the ToRs or the evaluators' interpretation.
Honduras	The report is not logically **structured** or user-friendly although the minimum content is present in the report. The **Executive Summary** is not a stand-alone section and does not provide a complete overview of the evaluation. The **methodology** section is very brief, does not explain methodological choice and justification in a clear manner, constraints

	or limitations are not addressed, and techniques and tools for data collection are nor provided in a detailed manner. Evaluation questions are not mentioned in this section. Some sources of quantitative and qualitative **data** have been identified but limitations are not clearly identified and there is no discussion of attempts to identify alternative data sources to improve credibility **of data** available. A very significant proportion of the **findings** section is descriptive rather than findings and analysis. **Analysis** is mostly based on results of a survey to national partners, UNFPA staff and UN agencies. Findings do not stem from rigorous data analysis and do not seem to be substantiated by evidence and cause and effect link between intervention and end results are not apparent. An evaluation matrix was not used to substantiate findings. Triangulation was not mentioned in methodology or applied to ensure sound analysis and credible findings. **Conclusions** are not based on credible findings or organized in priority order, and are for the most part findings and do not convey the evaluators' unbiased judgment of the interventions. **Recommendations** do not flow logically from conclusions, some flow from findings and assumptions. Recommendations for the most part are not strategic and targeted or presented in a priority order and are for the most part based on the survey conducted.
Iran	This is a **well-structured** and concise report that meets that needs stated in the **ToRs**. An excellent **Executive Summary** provides a clear overview of the report and is followed by a detailed explanation of **methodology**. Shortcomings in availability of **data** are discussed, as are the steps taken by the evaluators to compensate for any gaps. However the **analysis** is disappointing , limited to identifying where sub-outputs were achieved (due to lack of appropriate indicators or baseline data) rather than providing an evidence-based critique of programming choices. This undermines the subsequent **conclusions**, which have little information and are arranged by evaluation criteria and do not present a picture of the evaluators judgment. **Recommendations** are stronger but fewer and prioritized recommendations would have made them more effective.
Kyrgyzstan	The single narrative format of the **Executive Summary** does not provide a clear synopsis of the evaluation or make prominent the findings and **recommendations**. The lack of **methodological detail** or quantitative **data** reflects the paucity of evidence contained in the report, with the majority of **'Findings'** merely being descriptions of programme activities. There is a lack of detailed, evidence-supported discussion of **findings and analysis**. The inclusion of **conclusions** with findings is a flaw in the structure of the report and prevents the reader identifying any clear judgments by the evaluator. The evaluator attempts to reflect the broad ToRs in a **recommendations** section that totals over 52 recommendations which are unprioritised and lack a practical focus. The broad **ToRs** do not to provide the evaluator with sufficient focus which is essential given the extended duration of the Country Programme, but the evaluator did not revise them in light of logistical limitations.
Lao	The report's **structure** is clear, however, it includes an unconventional sequence of conclusions following recommendations. The clear **Executive Summary** is followed by a brief **methodology** section that states methodological activities rather than why these were selected. **Data** sources used are, in the main, reliable but not consistently referenced. The lack of a **conclusions** stage following on from **findings** results in too many recommendations and limits

	the scope of the evaluators to effectively communicate their judgment of interventions, and results in a lack of strategic **recommendations** for the CO.
Malawi	The scope of the evaluation is not sufficiently focused (due to too many evaluation questions and incorrect decision to assess CP's impact) and, as a result, the report does not provide detailed answers to a limited selection of useful questions. The report's analysis consists of a number of statements (classified under each outcome) which the reader may find difficult to relate to evidence. The **findings** do not stem logically from the analysis and the evaluators' interpretation -- expressed in very positive terms throughout the whole report -- does not appear well grounded or based on sound judgment. Although the design is described within a well structured **Methodological Approach** section, the **analysis** and findings are rarely supported by **data** and related sources of information. The Report does not provide a set of **conclusions**, but rather an overall final and very positive statement. This is followed by a list of numerous **recommendations**, some of which are useful, but are not prioritized (or presented within the **Executive Summary**) and are difficult to link to specific findings. Overall the report's positive tone is not substantiated by the reported analysis.
Mauritania	The short duration of the CPE stated in the **ToRs** is not sufficient for the detailed objectives and questions formulated by the CO. As a consequence, there is a lack of substance throughout the report. The **methodology** is the strongest section, including discussion of evaluability and justification of choice of methodological tools. However, this is undermined by key tools being missing from annexes. The **findings** are little more than a statement of activities, with little reference to **data** and what use of data can be found only serves to highlight the evidence gaps in many areas. Despite this, positive statements are made about the CP in the '**conclusion**' which confusingly located after the **recommendations**, and which themselves are a bullet point list of brief sentences that lack sufficient detail to be operationalized by the CO.
Mongolia	The scope of the evaluation is not sufficiently focused especially in view of the fact that: (i) this evaluation follows on the MTR which took place a year earlier; (ii) the poor quality of the CPAP P&T tool where baseline (and targets) were highly limited and indicators have not been tracked sufficiently to measure progress throughout the CP implementation. The **methodological design** (and its limitations) is well explained and the conduct of the evaluation is described in detail and evaluators' findings appear to be based on credible data and, in turn, support a rigorous **analysis** presented in a clear manner. Although **recommendations** stem logically from **findings** they are not presented in a priority order, nor are the main ones (solely) presented in an overly long **Executive Summary** which detracts from an otherwise detailed report. The report is hampered by a number of limitations including overly ambitious **ToRs** (which lack DAC criteria and should not have included assessment of CP impact).
Morocco	The report is clearly drafted but its **structure** does not meet the standard requirements expected from an evaluation report. The **Executive Summary**, which could have been slightly more concise, provides a good overview of the meta-evaluation results. The report addresses well the demands expressed in the **ToR**, in particular as regards the scope covered. However, the **methodological design** was insufficiently adapted to the specific challenges of a meta-evaluation, especially as far as the triangulation of **data** and information sources is concerned. This in turn puts into question the

	reliability of data used. The report is very descriptive, and findings are generally unsubstantiated. There is no part on **conclusions**, and **recommendations**, despite a visible effort to present them in a clustered and targeted way, suffer from this absence of logical linkage which should flow from findings to conclusions and to recommendations.
Mozambique	The report's **structure** is incomplete and not drafted according to the international standards (no **Executive Summary**, acronyms are not indicated in the text, **methodology** poorly explained, no indication of limitations). The report is written in a fragmented way, showing diverse styles and approaches which makes the report a sum of chapters with no coherence. No evaluation protocol is presented but rather a brief explanation of the process. Proper **analysis** of UNFPA contribution to results is not clearly presented, interventions are enumerated along with statements regarding national needs and contextual factors and cause and effect links between UNFPA support and results are very difficult to identify. Despite this, the lessons learned section contains useful observations; however it is not clear how these links to earlier analyses. Moreover, the lack of a proper **conclusions** stage following on from findings and results in too many **recommendations** and limits the scope of the evaluators to effectively communicate their assessment of results.
Myanmar	The report has a logical **structure** that begins with a clear **Executive Summary** that provides a useful snapshot of the evaluation. The **methodology** is described in detail, though the choice of methodology is not justified. The **Findings** section is lengthy and detailed, drawing on various types of evidence for analysis; however availability of **data** has created an imbalance in analysis of programme areas, and judgments are made by the evaluators that is not based on evidence or verified. **Conclusions** are weak and contained in a very brief section that bears little relation to the preceding findings section. Despite this shortcoming the recommendations are strategic and linked to **findings**, and limited in number but not prioritized. Overall, this is an effective report that meets the needs stated in the **ToRs** of analyzing evaluation findings and making clear and focused **recommendations** for the next CP, but has been undermined by weaknesses in key sections.
Panama	The report **structure** is comprehensive but not user-friendly, with gaps in some sections and a lack of logical structure in some places. The **Executive Summary** does not include a clear overview of the main results of the evaluation and lacks some key details. The **methodology** is described but briefly and is incomplete and methodological choice is not explained and justified. Triangulation is not mentioned or applied throughout the evaluation. Sources of qualitative and quantitative **data** have been identified, but limitations are not clearly identified or alternative credible data sources identified. A significant proportion of the findings and analysis section is solely descriptions of activities rather than findings and analysis. **Findings** do not stem from rigorous **data** analysis, are not clearly substantiated by evidence and are not presented in a clear manner. The **conclusions** do not clearly link to findings and are not presented in a priority order or clustered. **Recommendations** are not specific and practical and not linked to findings and **conclusions**. The evaluators have responded to the **ToRs** although deficiencies have not been highlighted by evaluation report.

Peru	The report is not well **structured**, but it is comprehensive in general. The **Executive Summary** provides an overview of the evaluation and is user-friendly but some sections are incomplete or weak. The **methodology**, including constraints and limitations, is not clearly explained or is missing and triangulation is not mentioned or systematically applied throughout the evaluation. There is no evaluation matrix displaying evaluation criteria, questions, **data** to be collected, methods, data analysis plan. Issues of data quality and credibility of primary and secondary data not discussed. Strategies to improve reliability and validity such as triangulating multiple sources of data is not mentioned or limitations to reliability of data are not made explicit. In some instances **findings** are substantiated by evidence; however some statements are not backed up by evidence, and in some cases sources are not identified. Cause and effect links are explained for interventions in each focus area for the most part. **Conclusions** are based on **findings** but could be strengthened by organizing them in priority order. The **Recommendations** are clear and practical but too general in some cases and are not presented in priority order, as well as being too numerous. Report responds to requirements in **TORs** but does not address weaknesses of **ToRs**.
Philippines	The report's **structure** includes all main elements but is too lengthy in key sections and the structure does not aid the reader in gaining a clear overview of the evaluation's results. Very little information or detail on the **methodology** is provided. Multiple sources of **data** are utilized but referencing is weak and there is low use of a number of other relevant evaluations. The **findings** section is lengthy in the description of activities but it is often unclear if this refers to UNFPA. There is little use of evidence to justify statements made about achievements. As a consequence, the **conclusions** struggle to synthesis the **findings** but do partially achieve this and are able to communicate the evaluators' judgments. **Recommendations** are poorly-written and lack sufficient specificity and detail to be of use to the CO. Overall the report is not well-integrated and does not meet the needs laid out in the **ToRs**.
Sao Tome and Principe	The CP evaluation lacks focus considering that the MTR took place a year earlier, and does not display any attempt by the evaluation team in the **methodology** to build on the detailed **ToRs** and identify an appropriate focus for the evaluation. The evaluation relies heavily on the CPAP tracking tool as a **data** source. The **findings** section is poorly presented, utilizing a confusing system of sub-sections that is more descriptive than analytical. The chosen structure of arranging **findings** by mandate fails to present a cohesive picture. **Recommendations** are not presented separately but listed at the end of each programme activity and often lack sufficient detail or practicality as well as being too numerous and unprioritised. There is an absence of any detailed **conclusions**. Overall this report fails to meet the requirements of the **ToRs**, mainly through the evaluation team's focus on description rather than analysis.
Senegal	Although the report is well written and provides us with a concise and clear **Executive Summary**, it suffers from a weak **structure**, which generates confusion between context description, findings and conclusions. The **methodological design** is not sufficiently explained and seems not fully appropriate to a context of scarcity of data. This scarcity of **data** and the absence of a proper analysis largely account for the limited credibility of **findings** when dealing with the results of the

	intervention. The section on **conclusions** fails to provide the reader with a clear set of organised and prioritised conclusions, logically stemming from findings. The section on **recommendations** also suffers from the absence of prioritisation and of options of operationalisation.
Syria	The **Executive Summary** is too long and prevents a useful overview of the evaluation being provided. **Methodology** is limited to descriptions of methods used, rather than a justification of choice, although there is a discussion of limitations and how they are addressed. There is limited use of reliable **data**, with evaluators relying on implementation of activities due to the absence of available data on the outcomes and results, and statements of findings are made without reference to evidence. **Findings** and Conclusions are combined in a single chapter, which opens with summary of conclusions per mandate which is unconventional and does not aid the reader. This lack of a separate set of **conclusions** makes it difficult to identify the logical flow from findings to conclusions to recommendations. The **recommendations**, which are in some cases operationally feasible, are too numerous for the CO to practically implement. The **ToRs** are detailed but include highly specific questions whose narrow scope has not been answered by the evaluators.
Tanzania	The **methodological design** appears justified and based on credible **data**, however, due to the short description in the report no definite judgment can be made based on the information provided in the report. **Findings** are outlined in a clear and logical way following an analysis that according to its presentation in the report can be considered sound. **Conclusions** and **recommendations** are very well structured and clearly refer to the findings. Unfortunately, the **Executive Summary** is too lengthy as to provide the reader with a short and concise overview.
Thailand	The report has a conventional **structure** with a brief and balanced **Executive Summary**. This is followed by an excellent introduction that provides a good level of detail on **methodology**, with clear explanation of tools. Detailed use of various **data** sources enabled robust and credible analysis with the production of detailed findings. However, this is not a consistent approach by the evaluators who, at times, make positive judgments about interventions that are not supported by evidence or rely too heavily on the views of implementing partners. There is then a clear weakness in how those findings that were robust were translated into **conclusions** by the evaluation team, with a confused and unclear chapter, which was combined with **recommendations**, which fails to deliver a clear message to the CO on the performance of its CP. Due to the weakness of this section, the report does not meet the needs of the evaluation in terms of producing practicable **recommendations** for the next CP**.**
Ukraine	The report's **structure** is not arranged in a clear manner, which is reflected in the **Executive Summary** that lacks key elements and fails to provide a clear overview or function as a standalone document. The **methodology** is presented in a disjointed way and there is little justification of selection of tools, how limitations were addressed or use of triangulation. **Data** shortages are made clear but there is little apparent effort to address gaps in reliability of data. **Findings** are not based on evidence and amount to little more than descriptions of activities or interventions by the CO. Conclusions are weak, and do not relate to findings, and in some cases refer to information not mentioned elsewhere in the report. **Recommendations** are too numerous and lengthy, with apparently important recommendations presented near the end

	of the section and not prioritized. Deficiencies in the **ToRs** are not commented on by the evaluators**.**
Yemen	The report is logically **structured** but lacks the clear separation of a conclusions section. The **Executive Summary** reflects the structure of the report with a focus on findings but does not include conclusions and recommendations. The report is based on a sound **methodological** design and utilizes **data** that are credible**.** However **findings** are not evidence-based. **Conclusions** – though linked to findings – are not well elaborated and **recommendations** based on **conclusions** could have been more detailed and would then have been of greater use to guide stakeholders on how best to strategize for the next country programme. The report responds to all evaluation criteria set out in the **ToRs** and the majority of evaluation questions.
Zimbabwe	The scope of the evaluation is not sufficiently focused due to too many evaluation questions outlined in the **ToRs** that the report did not cluster well enough to provide answers to the most useful questions. Instead the evaluation report reads as a detailed description of CP achievements and external factors but is not characterized by analytical depth. The **methodological design** is not described. **Findings** are in most cases not substantiated by evidence, and **conclusions** derived without foregoing analysis. Because no overall conclusions are presented it is difficult to understand if evaluation objectives have been met. **Recommendations** are targeted and limited in number but due to the weak nature of the conclusions they are based on not very strong. The **Executive Summary** is brief and well-structured but reflects the report's short coming of clearly outlining the main results of the evaluation.

Annex 5. CPE Survey questionnaire

1. Select Country
2. Select Region
3. Name of Person Completing Survey
4. Title of person completing survey
5. Please enter Country Programme Budget as it appears in the CPAP/CPD for the cycle that was evaluated (please enter amount in US$)
6. Please indicate total cost of the most recent Country Programme Evaluation (enter amount in US$)
7. Was the Country Programme Evaluation budgeted at the beginning of the cycle?
   Yes, No
8. Please enter the Atlas Budget code for the Country Programme Evaluation
9. Please indicate the starting date of the Country Programme Evaluation process as mentioned in TOR (Month/Day/Year)
10. Please indicate the end date of the Country Programme Evaluation process as mentioned in TOR (Month/Day/Year)
11. If actual start date of evaluation is different from the one mentioned in TOR, please indicate actual start date of the evaluation process (from drafting of TOR). Month/Day/Year
12. If actual end date of evaluation is different from the one mentioned in TOR, please indicate actual end date of the evaluation process (validation of final report). Month/Day/Year
13. Who drafted the ToR for the Country Programme Evaluation? Please select one or more options: CO M&E Officer/Focal Point, Consultants, Deputy Representative, Don't Know, Other.
14. Was the preparation of the TOR for the Country Programme Evaluation a consultative process involving country counterparts? Yes, No
15. Was any other UNFPA unit in HQ involved in the preparation of the TOR for the evaluation? Yes, No
16. Was any staff at UNFPA regional office involved in the preparation of the TOR for the Country Programme Evaluation? Yes, No.
17. Who was the evaluation manager in CO for the Country Programme Evaluation? Resident Representative, Deputy Representative, M&E Officer/Focal Point, Other, please specify.
18. Was a Reference Group created for the Country Programme Evaluation? Yes, No
19. Which stakeholders (organizations) were part of the Reference Group for the Country Programme Evaluation? Please select one or more options. Government, NGO, Academia, Civil Society Organization, Other UN Agency, Other, Please specify.
20. For what purpose(s) did the Reference Group meet? Please select one or more options. Draft/Review of TOR, Review/confirm evaluation team including task managers and consultants, Assist/Review data collection process, Review final evaluation report, Other please specify.
21. What was the composition of the core evaluation team? Please enter number of national consultants, international consultants, CO staff, others as applicable
22. Please indicate the duration of each consultant's contract
23. Please rate the availability of national expertise in the area of Reproductive Health in the Country. Did you encountered difficulties to recruit local evaluators? Not available, Limited Availability, Available, Easily Available.

24. Please rate the availability of national expertise in the area of Population and Development in the Country. Did you encountered difficulties to recruit local evaluators? Not available, Limited availability, Available, Easily Available.

25. Please rate the availability of national expertise in the area of Gender Equality in the Country. Did you encountered difficulties to recruit local evaluators? Not available, Limited availability, available, Easily available.

26. Please rate the availability of national expertise in the area of Evaluation Methodology in the Country. Did you encountered difficulties to recruit local evaluators? Yes, No, Don't know, Other please specify.

27. Were some of the evaluators (consultants) involved with the country programme (design and/ or implementation) under evaluation? Yes, No, Don't know, Other please specify.

28. Was an inception or desk report foreseen in ToR for the country programme evaluation? An inception report was foreseen, A desk report was foreseen, Both an inception and desk reports were foreseen, None were foreseen.

29. What was the procedure for the validation of the Final report? Please select one or more options. Approval by Reference Group, Approval by CO Representative with Regional Office Involvement, Approval by CO Representative, Other please specify.

30. Did the Country Office undertake a mid-term review of the country programme? Yes, No.

31. Please indicate the total cost of the mid-term review for the country programme (amount in US$)

32. During the course of the cycle corresponding to the country programme evaluation, how many evaluations (or reviews) were undertaken by the CO? Indicate number excluding CP evaluation and mid-term review:

33. What was the overall cost of all those evaluations (excluding CP evaluation and mid-term review) undertaken by the CO during the present cycle? Please enter amount in US$:

34. During the course of the cycle corresponding to the country programme evaluation, what types of evaluations (or reviews) were undertaken by the CO? Please select one or more options. Project evaluation, Focus Area Evaluation, Outcome Evaluation, Pilot Project, Other please specify.

35. Do you have a M&E Officer in your CO? Yes, No.

36. Do you have a M&E Focal Point in your CO? Yes, No.

37. If you have an M&E Officer or an M&E Focal Point, what was the nature of his/her involvement in the country programme evaluation? (please describe role and responsibility). Drafting Tor, Facilitating evaluation process, Management of evaluation process, Quality Control, Don't know, Other please specify.

38. Was the quality of the country programme evaluation assessed by your CO and/or Regional Office? By CO, By Regional Office, By both, No, Don't know, Other please specify.

39. If you answered yes to the previous question, was the quality of the evaluation assessed on the basis of a standardized quality assessment template? Yes, No.

40. Is there a management response sheet/ follow-up form available in your CO/ Regional Office? Yes, No.

41. If you answered yes to the previous question, who is responsible for the management response sheet/ follow-up form? Please indicate title. Resident Representative, Deputy Representative, M&E Officer/Focal Point, Don't know, Other please specify.

Annex 6. List of countries which participated in the CPE Survey

Algeria
Bangladesh
Brazil
Burkina Faso
Central African Republic
Chad
Dominican Republic
Egypt
El Salvador
English-Dutch Speaking Caribbean Countries
Eritrea
Gabon
Gambia
Ghana
Guinea
Haiti
Honduras
Iran
Kyrgyzstan
Lao People's Democratic Republic
Malawi
Mauritania
Mongolia
Morocco
Mozambique
Myanmar
Panama
Peru
Philippines
Republic of Yemen
Sao Tome and Principe
Senegal
Swaziland
Syrian Arab Republic
Tanzania
Thailand
Tunisia
Ukraine
Zambia
Zimbabwe