

STRENGTHENING EVALUATION FOR IMPROVED PROGRAMMING: UNFPA Evaluation Quality Assessment

STRENGTHENING EVALUATION FOR IMPROVED PROGRAMMING: UNFPA Evaluation Quality Assessment



United Nations Population Fund (UNFPA)
Division for Oversight Services
220 East 42nd Street
New York, NY 10017
www.unfpa.org



**Evaluation Report
Number 21**



Evaluation Report #21

.....

STRENGTHENING EVALUATION FOR IMPROVED PROGRAMMING: UNFPA Evaluation Quality Assessment

NOTE: The views expressed in this report do not necessarily reflect those of the United Nations Population Fund.



Tony Beck (independent),
Francoise Coupal (independent),
and Scott Green (DOS),
with Christina Bierring (DOS)

DECEMBER 2005

Foreword

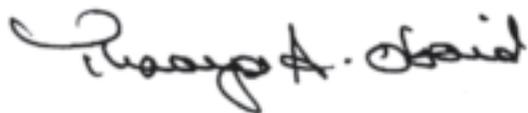
Evaluation is an important management tool that enables us to make informed decisions based on learning. It is a key element of Results-Based Management and is well integrated in UNFPA's programming and strategic planning processes. As UNFPA works in an ever-changing environment, not only within the UN system but also in the broader environment of development aid and national context, it is important that the Fund uses this tool effectively so that it remains relevant and meaningful. Evaluation can tell us if we are doing the right thing in the right way, and how we can improve our work to ensure that UNFPA's activities and programmes have greater impact. Thus, evaluation helps us to develop a stronger evidence base to support UNFPA's work. It also helps the organization to be more accountable for results; this contributes to empowering our ultimate beneficiaries.

In 2005, UNFPA commissioned the first-ever meta evaluation – evaluation of evaluations. We wanted to have a hard look at the quality of our evaluation work, using recognized international standards as a benchmark. We also wanted to determine what kinds of improvements were needed.

The Evaluation Quality Assessment tells us clearly that UNFPA needs to improve the quality of its evaluations. It is increasingly important to pay greater attention to using standard evaluation criteria, such as relevance, effectiveness, efficiency, impact and sustainability, and to ensure that all evaluations provide an assessment of the organization's interventions from a gender mainstreaming and human rights perspective.

This meta evaluation has provided a number of valuable recommendations for action; we have included a detailed management response. As we move forward with our follow-up to the report's findings and recommendations, UNFPA will strengthen its evaluation capacity as well as practices. We further plan to conduct a second meta evaluation in 2010 in order to determine whether our new measures have brought about any measurable improvements.

I am confident that our renewed efforts to improve the quality of evaluation will lead to better use of evaluation findings and recommendations and, ultimately, to improving the lives of people we serve.



Thoraya Ahmed Obaid
Executive Director
United Nations Population Fund (UNFPA)

Table of Contents

Acknowledgements	6
Acronyms	7
Executive Summary	8
1. Background	18
1.1 Introduction and objectives of the meta-evaluation	18
1.2 Use and users	19
1.3 Methodology.....	19
1.3.1 Quantitative assessment.....	19
1.3.2 Qualitative assessment.....	20
1.4 Organization of the report	22
2. Main findings from the assessment of report	23
2.1 Overall quality of evaluation reports	23
2.1.1 Quality of evaluations as assessed against the 1997 Policies and Procedures Manual	25
2.1.2 Quality of evaluations as assessed against international criteria	25
2.2 Quality of attention to individual EQA criteria	26
2.2.1 Quality of terms of reference	26
2.2.2 Completeness of evaluations	27
2.2.3 Identification of use and users.....	27
2.2.4 Quality of methodologies employed	27
2.2.5 Is UNFPA evaluation relevant to its Strategic Framework?.....	28
2.2.6 Gender mainstreaming	32
2.2.7 Human rights.....	33
2.2.8 Quality of use of the OECD-DAC evaluation criteria	34
2.2.9 Quality of lessons learned.....	35
2.2.10 Recommendations and their follow-up	36
2.2.11 Evaluation of training.....	37
3. Quality of evaluation organisation and processes	39
3.1 The evaluation culture and approaches used at UNFPA	39
3.2 Who carries out UNFPA evaluations	42
3.3 Length of time devoted to evaluations	43
3.4 Cost-benefit analysis	44
3.5 Supporting national capacity.....	44
4. Quality of evaluation resources	46
5. Conclusions	47
6. Lessons learned	49
7. Recommendations and Management Response	50
Bibliography	54

ANNEXES

Annex 1 Meta-evaluation Terms of Reference	57
Annex 2 Sampling, assessment process, the EQA matrix and data analysis	68
Annex 3 List of 60 sample evaluations	91
Annex 4 Interview Guides	95
Annex 5 List of people interviewed	103
Annex 6 Background to team members	109
Annex 7 Bias and limitations	110

BOXES

Box 1 What is meta-evaluation?	18
Box 2 What Appreciative Inquiry is and why we used it.	21
Box 3 Having a seat at the table: leveraging evaluation within the policy dialogue.....	31
Box 4 Evaluating policy dialogue?	31
Box 5 Good practice in attention to gender equality: the cases of Sri Lanka and Mozambique...33	
Box 6 Building new partnerships for evaluation: the case of Vietnam	40
Box 7 'Seeing our footprints in the new Country Programme.'	41
A participatory evaluation in Iran	
Box 8 Selection of evaluators by the Lebanon Country Office.....	43

FIGURES

Figure 2.1 Summary of ratings for 60 evaluations on all criteria	23
Figure 2.2 Rating by 9 1997 criteria	25
Figure 2.3 Rating by 22 international criteria	25
Figure 2.4 Clear definition of use and users	27
Figure 2.5 Good practice in methodology	28
Figure 2.6 Rating against 6 UNFPA specific criteria.....	29
Figure 2.7 Quality of assessment of gender mainstreaming	32
Figure 2.8 Quality of assessment of human rights	34
Figure 2.9 Quality of use of the OECD-DAC criteria.....	34

TABLES

Table 2.1 Rating of 60 evaluation reports by criteria.....	24
Table 2.2 Average rating on six terms of reference criteria.....	26
Table 2.3 Rating of 60 evaluation reports by UNFPA specific criteria	29
Table 3.1 Profile of good practice case studies.....	39

Acknowledgements

We are particularly grateful to **Olivier Brasseur** (Director, Division for Oversight Services) and **Linda Sherry-Cloonan** (formerly Deputy Chief of the Office of Oversight and Evaluation, currently Deputy Director Division for Human Resources) for their extensive input and support.

The Reference Group established as a guide and sounding board should be thanked for strategic contributions throughout: **Esteban Caballero, Jacqueline Desbarats, Garimela Giridhar, Patricia Guzman, Nobuko Horibe, Janet Jackson, Nicola Jones, Steven Kraus, Petra Lantz, Ian Mcfarlane, Kemal Mustafa, Brendan O'Brien, Diego Palacios, Daniel Sala-Diakanda, Aminata Toure, and Mieko Yabuta.**

A large number of UNFPA staff at HQ and in Country Offices and CSTs, as well as UNFPA counterparts, submitted to interviews. We would like to thank in particular the **Country Office staff** who arranged our missions so efficiently and pleasantly **in Ghana, Lebanon, Mozambique, Nicaragua, Sri Lanka, and Viet Nam; CST staff in Addis Ababa, Bangkok and Mexico; and all beneficiaries, Government and NGO staff who participated.**

Colleagues at UNDP and UNICEF also provided key inputs. **Magalye Mars-Mompont** provided competent administrative support.

Thanks to all.

Acronyms

AI	Appreciative Inquiry
ALNAP	Active Learning Network on Accountability and Participation
CO	Country Office
COAR	Country Office Annual Report
CP	Country Programme
CST	Country Support Team
DALY	Disability Adjusted Life Year
DOS	Division for Oversight Services
EQA	Evaluation Quality Assessment
ICPD	International Conference on Population and Development
KAP	Knowledge, Attitude and Practice
MTR	Mid Term Review
MYFF	Multi-Year Funding Framework
OECD-DAC	Organisation for Economic Cooperation and Development - Development Assistance Committee
OVI	Objectively Verifiable Indicators
PM&E	Planning, monitoring and evaluation
PRSP	Poverty Reduction Strategy Paper
RBM	Results based management
RH	Reproductive Health
SWAps	Sector Wide Approaches
TSD	Technical Support Division
UNCT	United Nations Country Team
UNEG	United Nations Evaluation Group

Executive Summary

BACKGROUND AND OBJECTIVES

In order to review and improve the quality of UNFPA evaluations, the Division for Oversight Services conducted a meta-evaluation between January and December 2005, entitled Evaluation Quality Assessment (EQA). It had three short-term objectives:

- ❑ To facilitate organizational learning about effective evaluation practices.
- ❑ To highlight effective evaluation practices that can be replicated.
- ❑ To establish a baseline for future quality assessments.

It also had one longer-term objective:

- ❑ To produce more consistent and better quality evaluations.

This report is organized around four main themes: evaluation quality, evaluation practice, process and follow-up, and quality of evaluation support provided by UNFPA.

METHODOLOGY

The meta-evaluation adopted a mixed-method approach, triangulating between quantitative and qualitative techniques.

Quantitative techniques involved assessing a representative sample of 60 evaluations measured against 28 good practices criteria, on a five point rating scale. It also included a separate assessment of the terms of reference for carrying out evaluations. Quality of the sample was assessed against three main reference points – the 1997 Policies and Procedures Manual; international good practice; and current UNFPA programme priorities.

Qualitative techniques focused on the evaluation processes that led to the adoption of good practices. This process involved:

- Interviews with 38 UNFPA Headquarters and CST staff;
- The documentation of six good practice case studies spanning all four major regions – in Ghana, Lebanon, Mozambique, Nicaragua, Sri Lanka, and Viet Nam. These six case studies also involved interviews with four UN Resident Coordinators, 39 UNFPA Country Office staff, 52 counterpart staff in government and civil society, 18 evaluators and 16 beneficiaries;
- One TSD good practice case study.

The main interview techniques used were semi-structured interviews, either individually with senior staff, or in focus groups. As a central feature, these interviews included the use of Appreciative Inquiry techniques. Good practice examples are highlighted throughout the report.

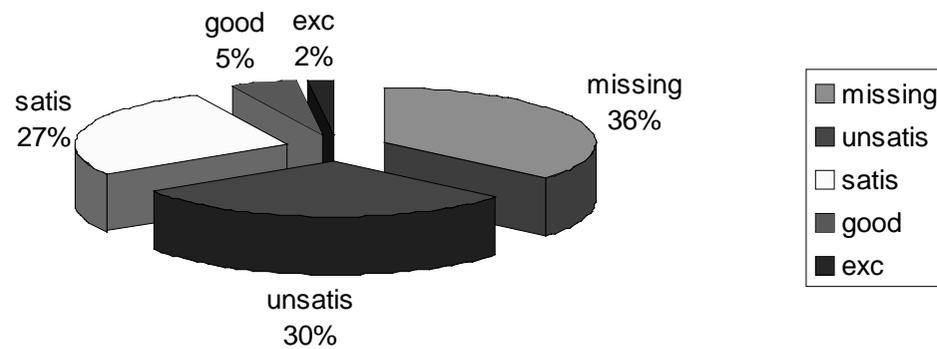
The meta-evaluation adopted a utilization-focused approach, based on intended use for intended users. This methodology was selected because the meta-evaluation concentrated on improving organizational learning, highlighting effective practices and improving quality. This in turn, requires substantial buy-in from UNFPA staff, which utilization focused approaches can help foster. An important part of the utilization-based approach was the establishment, at the start of the exercise, of a UNFPA Reference Group of potential users. The meta-evaluation process involved interactions between the meta-evaluation team and the Reference Group at key points throughout the process.



FINDINGS ON EVALUATION QUALITY

Just over one-third (34 per cent) of evaluations rated satisfactory or better across all criteria (Figure 2.1). This is unsatisfactory and requires improvement. A regional breakdown of rating results suggests that weak evaluation performance is an organizational issue. Senior management (e.g. the Executive Director, Deputy Executive Directors and Heads of Geographical Divisions) have reason to question whether they have adequate and trustworthy information on the results of programmes, either to demonstrate UNFPA performance, or to improve programmes. However, good practice in almost all evaluative areas could be identified only in a minority of cases.

Figure 2.1 Summary of ratings for 60 evaluations on all criteria



The assessment against the three main reference points used in this meta-evaluation is summarized below:

Reference point	Missing (%)	Unsatisfactory(%)	Satisfactory (%)	Good (%)	Excellent (%)
9 standards in the 1997 Policies and Procedures Manual	16	35	40	6	3
22 international good practice standards	32	30	31	5	2
6 UNFPA specific standards	18	44	29	6	3

Evaluation quality as assessed against the nine standards in the *1997 Policies and Procedures Manual* was higher than for the other two reference points, with 49 per cent of evaluations rated as satisfactory or better. However, it can be concluded that one in two of UNFPA evaluations did not meet policy requirements.

Rating across all criteria for the entire sample is given below.

RATING OF 60 EVALUATION REPORTS BY CRITERIA (PERCENTAGES)					
Criteria	Assessment				
	Missing	Unsatisfactory	Satisfactory	Good	Excellent
TOR*					
1.a. Purpose/Rationale of the evaluation is clearly defined.	64	4	28	5	0
1.b. Use and users clearly defined.	74	14	8	2	3
1.c. Scope and Focus is clearly identified	52	7	24	16	2
1.d. Methodology specified	63	25	11	2	0
1.e. Roles and responsibilities of evaluators defined	78	12	4	6	0
1.f. Clear Outputs, deadlines, formats specified	70	21	7	2	0
THE EVALUATION REPORT					
3. Quality of Executive Summary.	45	17	27	7	4
4. Purpose of the evaluation is outlined	35	19	38	7	1
5. Use and users clearly defined	58	30	8	4	0
6. Objectives of evaluation clearly defined	13	11	69	6	1
7. Quality of methodology					
7.a. Link to evaluation objectives/questions	10	17	68	4	1
7.b. Good practice	9	64	18	6	3
7.c. Bias and limitations	49	20	23	4	4
8. Consideration given to propriety and ethics	89	8	2	0	1
9. Evaluator values and bias outlined	91	7	2	0	0
10. The project/programme to be evaluated was clearly described	0	65	17	10	8
11. The role and contributions of UNFPA and other stakeholders were clearly defined	10	46	36	3	5
12. Extent of Stakeholders/beneficiaries involvement	82	10	5	3	0
13. Quality of assessment of gender mainstreaming	18	45	21	7	9
14. Quality of assessment of capacity development	5	37	51	6	2
15. Quality of assessment of human rights	67	24	9	1	0
16. Evaluation enables UNFPA to engage in evidence based policy dialogue	35	36	24	3	2
17. Quality of assessment of cultural issues	40	32	21	5	2
18. Quality of assessment of national ownership	5	45	42	7	1
19. Quality of assessment of partnership and alliance building	12	74	7	3	4
20. Quality of assessment of RBM	13	47	22	12	6
21. Quality of assessment with regards to relevance	3	34	54	6	3
22. Quality of assessment with regards to effectiveness	0	51	36	8	5
23. Quality of assessment with regards to efficiency	58	34	7	1	0
24. Quality of assessment with regards to sustainability	20	32	39	8	1
25. Quality of assessment with regards to impact	21	38	33	4	4
26. Quality of conclusions	8	36	47	7	2
27. Quality of recommendations	2	20	71	7	0
28. Quality of lessons learned [^]	-	35	52	11	2

*Based on 28 evaluations which included terms of reference.

[^]Based on 33 evaluations which included a section on lessons learned. 'Missing' was not included for this criteria.

The meta-evaluation had the following findings concerning individual criteria:

Terms of reference for the evaluation were rated higher than the evaluations themselves, with 41 per cent scoring satisfactory or better. The terms of reference were rated higher on defining the purpose/rationale, and scope and focus of the evaluation; and lower on identifying use and users, and specifying the methodology to be used. The terms of reference that clearly define the evaluation purpose, users and uses, as well as evaluation objectives and methodology, is one key element that will lead to better quality evaluations. For 11 evaluations where the terms of reference received a rating of 8 or more, the average rating for the report on criteria 3-28 was 1.26. For the remaining 20 evaluations containing terms of reference, the average rating for the evaluation on criteria 3-28 was 1.14.¹ Hence, it is apparent that higher quality terms of reference is one factor contributing to a higher quality of evaluation.

Completeness of reports. About half of the evaluations included an Executive Summary, terms of reference, bibliography, and list of people met. Data collection instruments were included in 32 per cent of cases.

Lack of attention to defining use and users is a major gap and likely to hinder follow-up to recommendations.

Methodological description was skimpy and opaque, with few details on sampling and limited emphasis on triangulation. In terms of good methodological practice, 27 per cent of evaluations were rated as satisfactory or better. Two distinct methodological weaknesses appeared: lack of attention to gender equality, and marginalized populations. Furthermore, UNFPA evaluations rarely canvassed the opinions of poorer groups, and tended to be non-participatory. Evaluations with higher rated method sections also scored better for the evaluation as a whole. The 13 evaluations which rated satisfactory or better on good practices in methodology scored an average of 1.24 for the report as a whole, as opposed to 1.11 for the remaining 47 evaluations.

Attention to gender in the method section also led to greater attention to gender in the evaluation report. Of the 13 evaluations which specifically addressed gender in their method section, 54 per cent were rated satisfactory or better on the criteria of 'quality of assessment of gender mainstreaming', as opposed to 33 per cent for the remaining 47 evaluations.

Across the **six specific criteria**, assessment of capacity development was rated highest, with 59 per cent satisfactory or better, and assessment of partnership and alliance building lowest, with 14 per cent rating satisfactory or better. Attention to cultural issues was also below the average for the sample as a whole, with only 28 per cent rating satisfactory or better. Assessment of results-based management (RBM), national ownership and the extent to which the evaluation is used in policy dialogue were rated roughly equivalent to the sample as a whole.

Given the current focus of evaluations at the project and output levels, there is limited information on **evaluating policy dialogue and advocacy**. Twenty two of the 25 respondents questioned noted that as UNFPA moves deeper into Sector-Wide Approaches (SWAs) and general budget support, evaluations will need to pay increased attention to the results of policy dialogue.

Nearly two-thirds (63 per cent) of evaluations were rated as unsatisfactory or missing on **gender mainstreaming**. This is a troubling finding given that one of the Multi-Year Funding Framework (MYFF) 2004-2007 goals is the promotion of gender equality and empowerment of women. It means that UNFPA has limited evaluative information on gender mainstreaming to support improved programming.

¹ Data was analysed as follows. An average score was calculated on criteria 3 to 28 for the sample as a whole, with no weighting allocated to criteria, and missing=0, unsatisfactory=1, satisfactory=2, good=3, excellent=4. A similar average was calculated for terms of reference.

Ten per cent of evaluations achieved a satisfactory or better rating on **assessment of human rights**, but identified no good practices. In 66 per cent of cases there was no reference to human rights.

Use of the **OECD-DAC criteria** was one of the strengths of the sample, in particular use of relevance and effectiveness. Overall 57 per cent of evaluations were rated as satisfactory or higher on two or more of the DAC criteria. Because evaluations tended to focus at the output level, impact and sustainability were not covered well. Attention to efficiency was the lowest rated of the OECD-DAC criteria. Lack of attention to coverage, in particular of poorer groups, was a major gap. There was limited information in the evaluations on which sections of the population UNFPA funding was able to reach.

In terms of **quality of recommendations**, 71 per cent of UNFPA evaluations were rated as satisfactory, and 7 per cent as good. However, it is likely that given the overall unsatisfactory quality of UNFPA evaluations, there is limited follow-up. This is also the message we received from semi-structured interviews. Of the respondents questioned in relation to follow-up, 41 thought there was inadequate follow-up, while only 21 thought follow-up was adequate. Furthermore, 19 of the 21 positive responses were from evaluators or government staff involved in good practice case studies, and were referring specifically to the good practice evaluation. No one at Headquarters thought there was adequate follow-up to evaluation recommendations.

In the good practice case studies there has been extensive follow-up to recommendations, in particular where the evaluators adopted a participatory approach, such as in Lebanon, Mozambique, and Viet Nam. Country Offices have also been experimenting with recommendation follow-up matrices, which are increasingly becoming standard in evaluation practice.

Evaluating training exercises was an important element in a majority of reports. Two-thirds (66 per cent) of the sample included a discussion of training programmes; training was the main evaluation focus in one quarter of all samples. However, the main focus of the evaluation of training was on reactions of trainees, rather than on documenting learning or behavioral changes, even though some good evaluation practices were identified.

THE EVALUATION CULTURE AT UNFPA

There is no disagreement at Headquarters or in the field of the importance of strengthening the evaluation function at UNFPA. On the positive side, Country Offices have been innovative, producing useful evaluations which have helped to reshape programming, making it more effective. This general support throughout the organization, and the ability of UNFPA to achieve good practices across a range of Country Offices, suggests there is a solid base on which to build better evaluation practices.

However, our finding from interviews, particularly at Headquarters, was that UNFPA currently lacks a strong evaluation culture, and evaluations do not play a strategic role in decision-making. Evaluation is not perceived by most staff interviewed as having a coherent function at UNFPA, suggesting a need for an improved organizational evaluation strategy.

Both DOS and country level evaluations were perceived as being too “accountability focused”, and too threatening, suggesting the need for more focus on participatory and lesson learning models. Respondents told us there is a culture of blame, rather than a culture of learning. A total of 78 out of 108 respondents noted that evaluations should be more participatory, with no contrary views. Respondents said they need better guidance on what kinds of evaluation models are relevant in different contexts, and the benefits and drawbacks of different approaches. CSTs were also uncertain as to their role regarding evaluation. These findings explain some of the reasons for low scoring against the EQA matrix.

What UNFPA can achieve in terms of evaluation quality is partly dependent on the host government. Where there is sufficient government capacity and willingness to participate, as in the cases of Sri Lanka, Viet Nam and Lebanon, evaluation practice is significantly stronger.

WHO CARRIES OUT UNFPA EVALUATIONS

Fifty four of the evaluations included information on the institutional background of the evaluators. Of this sub-sample, 46 per cent were carried out solely by national evaluators, a further 30 per cent by national and international evaluators, with the remaining 24 per cent performed by international evaluators, CST advisers, or a combination of these. Given the preponderance of national evaluators, individually or in combination with international evaluators (76 per cent of cases), a key factor for improving the quality of UNFPA evaluations is for Country Offices to utilize the services of qualified national evaluators.

Higher quality evaluations appear to be produced by mixed teams of national and international evaluators. The six evaluations involving mixed teams of national and international evaluators (excluding CST advisers) received a 50 per cent satisfactory or better rating, as opposed to 34 per cent for the sample as a whole. This was not always the case, however. In Lebanon, Viet Nam and Iran, for instance, we found high quality evaluations conducted solely by skilled and experienced national evaluators.

CST advisers were involved in 22 per cent of evaluations. Our visits to CSTs suggest that while CST advisers have strong technical backgrounds they do not always have the requisite skills to conduct evaluations, have not been trained recently on evaluation, and/or are available for very short periods of time.

TIME SPENT ON EVALUATIONS

The average length of time devoted to evaluations was 20 days (based on a representative sub-sample of 17 evaluations). While evaluation quality depends on a number of factors, there appears to be a strong correlation between length of time spent on the evaluation and its quality: the average length of time spent on the 11 evaluations included in the good practice case studies was 42 days. In the 28 evaluations which included a section on 'bias and limitations', lack of time was mentioned as a constraint in 20 of them, or 71 per cent of cases. In addition, 17 evaluators interviewed (out of the sample of 18) cited lack of time as a major constraint to producing good quality evaluations.

COST-BENEFIT ANALYSIS

The meta-evaluation made a general assessment of costs and benefits. In five out of seven good practice case studies savings (or potential savings) in programming were higher than the cost of evaluation.

SUPPORTING NATIONAL CAPACITY

Improving national capacity will likely lead to an improvement in the quality of UNFPA evaluations. We did find some evidence in the case study countries of attempts to improve national evaluation capacity. Staff at both Country Office and Headquarters levels felt that the central role of the UN should be to support the development of national capacity. Some 57 interviewees were asked questions concerning national capacity building. Of these, 31 respondents thought that it should be a future priority, with no contrary views.

QUALITY OF EVALUATION RESOURCES

Country Office respondents were unanimous in their praise for the Headquarters-led Regional Dialogues on Results-based Planning, Monitoring and Evaluation. Other HQ supported initiatives received mixed reviews. It was found that the Policies and Procedure Manual was not being used to guide evaluations. In addition, the Programme Manager's Planning, Monitoring and Evaluation Toolkit was not being used on a regular basis, and respondents requested more operational content and an easier to use format. Of the 64 interviewees who asked questions on HQ resources, 31 agreed with the comments above, with three holding contrary views.

RECOMMENDATIONS

On the basis of comments received from the Reference Group and other respondents, recommendations are presented in order of priority. UNFPA is urged to produce a management response to these recommendations, noting: timelines, budgets, responsibilities, and monitoring and assessment of implementation.

Improving evaluation quality

FINDINGS	RECOMMENDATIONS
<p>Sixty six per cent of the evaluation sample rated unsatisfactory or missing on the EQA matrix. Evaluation quality at UNFPA is currently unsatisfactory and requires improvement (Section 2.1).</p> <p>UNFPA currently lacks a strong evaluation culture, and evaluations do not have a strategic role in decision-making. Evaluation is not perceived by staff interviewed as having a coherent function at UNFPA, and evaluation findings are not fully utilized (Section 3.1).</p> <p>Regarding questions for improving evaluation quality at UNFPA, 78 out of 108 noted that evaluations should be more participatory. DOS and country level evaluations were perceived as being too "accountability focused", and too threatening. Good practice cases revealed that where there is greater participation of country counterparts in the evaluation, there is also more consistent follow-up to evaluation recommendations (Section 3.1).</p> <p>Current time spent on evaluations is too short. The average length of time spent on the 11 evaluations included in the good practice case studies was 42 days, as opposed to 20 days for 17 other representative evaluations which included data on timing (Section 3.3).</p>	<p>1. UNFPA should develop, implement, and monitor a strategy for improving the quality of evaluations.</p> <p>This could include:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Updating the Policies and Procedures Manual section on monitoring and evaluation, following UNEG guidance.² <input type="checkbox"/> Determining what levels of improvement in evaluation quality is required, and over what time period. Meeting these targets could be measured by a full-scale meta-evaluation, similar to the current one, in 2010. There could also be periodic smaller-scale assessments of quality - for instance at six month intervals. <input type="checkbox"/> Providing Country Offices and CSTs with training in participatory and utilization focused evaluation approaches. This could highlight different entry points for participation (e.g. development of ToRs, workshops on results); and different methods that can be used to promote participation of stakeholders, including counterparts and beneficiaries. <input type="checkbox"/> Hiring additional monitoring and evaluation staff, to be located at the regional (e.g. CST) or Country Office level. However, UNFPA should examine the experience of establishing focal points for particular sectors before deciding on this issue. If new staff are hired as PM&E focal points, this should not be a substitute for mainstreaming PM&E throughout the organisation.

2. The UN Evaluation Group Norms for Evaluation in the UN System (April 29 2005).

Improving evaluation quality *(continued)*

<p>Country Office respondents were unanimous in their praise for the HQ led Regional Dialogues on Results-based Planning, Monitoring and Evaluation (Section 4).</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Integrating monitoring and evaluation in core competencies of relevant staff. For example, evaluation could be included in the Analytic and Strategic Thinking, and Results Orientation competencies. <input type="checkbox"/> The EQA matrix could be adapted to a set of quality standards that could be attached to the evaluation terms of reference. <input type="checkbox"/> Providing guidance on the optimal length of time for different kinds of evaluations. <input type="checkbox"/> Continuing to invest in Regional Dialogues. <input type="checkbox"/> Maintaining a searchable electronic database of evaluations. This would require screening of evaluations against key criteria in the EQA assessment matrix to determine acceptance in the database (e.g. completeness of report, quality of methodology, attention to key cross-cutting themes, and OECD-DAC criteria). The database should be staffed with a full-time manager, as inadequately managed databases are not used. The data base manager could also support future meta-evaluation activity.
--	---

Follow-up recommendations

FINDINGS	RECOMMENDATIONS
<p>There is limited follow-up to evaluation recommendations (Section 2.2.10). In the good practice case studies there had been extensive follow-up to recommendations, in particular where the evaluators had adopted a participatory approach, such as in Lebanon, Mozambique, and Viet Nam.</p> <p>Eighty eight per cent of evaluations were rated unsatisfactory or missing in terms of clear identification of use and users (Section 2.2.3).</p>	<p>2. All evaluations should include an action plan for follow-up to recommendations.</p> <p>This should include response to recommendations by relevant parties, prioritization of recommendations, and a recommendations tracking matrix. COARs should report on management attention to follow-up on recommendations.</p> <p>Evaluations should be expected to identify clearly use and users in terms of reference and reports.</p> <p>The evaluation Executive Summary should be translated into local languages.</p>

Evaluating current strategic priorities

FINDINGS	RECOMMENDATIONS
<p>An analysis of six UNFPA specific criteria - assessment of capacity development, cultural issues, national ownership, partnership and alliance building, and the extent to which the evaluation enables UNFPA to engage in evidence based policy dialogue - revealed that attention to these criteria was satisfactory or better in 38 per cent of cases. Currently, UNFPA is not capable of answering evaluative questions posed by the Strategic Direction Positioning Statement (Section 2.2.5).</p>	<p>3. Guidance and staff capacity development is needed for evaluation of current strategic priorities.</p> <p>This should include guidance on evaluating participation in Sector Wide Approaches.</p>

Gender mainstreaming and human rights

FINDINGS	RECOMMENDATIONS
<p>Sixty three per cent of evaluations were rated as unsatisfactory or missing on gender mainstreaming (Section 2.2.6).</p> <p>Ten per cent of evaluations achieved a satisfactory or better rating on quality of attention to human rights, with no good practice identified. In 66 per cent of cases there was no reference to human rights (Section 2.2.7).</p> <p>Attention to gender in the evaluation method section does lead to greater attention to gender in the evaluation report. Of the 13 evaluations which specifically address gender in the method section, 54 per cent were rated satisfactory or better on the criteria of 'quality of assessment of gender mainstreaming', as opposed to 33 per cent for the remaining 47 evaluations (Section 2.2.4).</p>	<p>4. UNFPA needs to demonstrate a measurable improvement in attention to human rights and gender equality in its evaluations. It can do this by, inter alia:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Ensuring that terms of reference include promoting human rights and gender equality as areas to assess. <input type="checkbox"/> Ensuring that methodologies include details on how contributions to human rights and gender equality will be measured. <input type="checkbox"/> Ensuring that data in evaluation reports is sex-disaggregated. <input type="checkbox"/> Locating evaluators with a good understanding of the importance of human rights and gender mainstreaming. <input type="checkbox"/> Ensuring that there are separate sections on human rights and gender equality in the evaluation, and that they are also covered in all other main sections of the evaluation as a cross-cutting theme. <input type="checkbox"/> Ensuring that at least one recommendation deals directly with human rights, and one with gender equality.

National capacity

FINDINGS	RECOMMENDATIONS
<p>Forty six per cent of evaluations in the sample were carried out solely by national experts and a further 30 per cent by a mixed team of national and international evaluators. A key factor for improving the quality of UNFPA evaluations is for Country Offices to access high quality national evaluators (Section 3.2).</p> <p>Higher quality evaluations appear to be produced by mixed teams of national and international evaluators (Section 3.2).</p>	<p>5. A long-term strategy for UNFPA is to support, with other UN agencies, national evaluation capacity.</p> <p>6. Mixed teams of national and international evaluators should be used where feasible.</p>

Methodologies

FINDINGS	RECOMMENDATIONS
<p>In terms of good practice in methodology, 27 per cent of evaluations rated satisfactory or better, and 64 per cent unsatisfactory. Nine per cent of the sample did not include any methodological details at all (Section 2.2.4).</p> <p>Evaluations which rated satisfactory or better on good practice in methodology were also higher quality evaluations. Greater attention to method will likely lead to a stronger report (Section 2.2.4).</p>	<p>7. Planned method sections of evaluations could be screened (e.g. against an adapted set of EQA standards - see Recommendation 1). Evaluations over a certain budget limit, and/or a random sample of evaluations, could have a methodological review outside of the Country Office.</p>

Completeness of reports

FINDINGS	RECOMMENDATIONS
<p>Overall, terms of reference were unsatisfactory. Where terms of reference were rated higher, evaluations also received a higher rating (Section 2.2.1).</p> <p>Fifty two per cent of the evaluations included an Executive Summary, 48 per cent included a bibliography, and 54 per cent included a list of people met. Data collection instruments were included in 32 per cent of cases. Terms of reference were included in 46 per cent of cases. (Sections 2.2.1 and 2.2.2).</p>	<p>8. Screening of terms of reference could be carried out by Geographical Divisions, with spot-checking by Division for Oversight Services (DOS). A quality checklist for terms of reference could be developed, building on the guidance in the Programme Manager's Planning, Monitoring and Evaluation Toolkit.</p> <p>9. A checklist based on the model Table of Contents in the Programme Manager's Planning, Monitoring and Evaluation Toolkit could be developed for screening of Tables of Contents.</p>

HQ guidance

FINDINGS	RECOMMENDATIONS
<p>The Programme Manager's Planning, Monitoring and Evaluation Toolkit was not being used on a regular basis, and respondents requested more operational content, and an easier to use format. (Section 4).</p>	<p>10. HQ resources should be edited to make them more user-friendly. A professional editor should be hired for this purpose.</p>

1. Background

1.1 INTRODUCTION AND OBJECTIVES OF THE META-EVALUATION

UNFPA produces a substantial body of evaluation work each year. It is intended to support the development and improvement of its programmes, and underscore accountability for achieving results. The evaluation function at UNFPA is decentralized, meaning that almost all evaluations are carried out by Country Offices. Country Offices produce about 120 evaluation reports every year including reviews, at a cost of some US\$3 million. UNFPA also commissions a number of regional and inter-regional evaluations. Consequently, the current investment in evaluation is substantial.

To review and improve the quality of UNFPA evaluations, DOS conducted a meta-evaluation between January and December 2005. This is in line with current thinking in the evaluation field on how to improve evaluation quality, and actual practice by several donors, UN agencies and aid networks. The UNFPA meta-evaluation was titled: Evaluation Quality Assessment (EQA). It had three short-term objectives:

- ❑ To support organizational learning about effective evaluation practice.
- ❑ To highlight effective evaluation practices that can be replicated.
- ❑ To establish a baseline for future quality assessments.

It also had one longer-term objective:

- ❑ To produce more consistent and better quality evaluations.

BOX 1: WHAT IS META-EVALUATION?

Meta-evaluation usually involves two parts:

- **An assessment of the quality of a representative sample of evaluation reports, measured against a set of good practice standards.**
- **A synthesis of evaluation findings.**

Meta-evaluations are intended to improve evaluation practice and quality, which should then help to improve programming.

This report is organized around four themes: evaluation quality, evaluation practice, process and follow-up and quality of evaluation resources provided by UNFPA. Further background details can be found in the EQA Terms of Reference in Annex 1.

UNFPA is also carrying out a synthesis of evaluation findings parallel to its investigation of evaluation quality, which will be published separately.

1.2 USE AND USERS

The main users are intended to be UNFPA staff and counterparts involved in planning, monitoring and evaluation (PM&E). The main uses, as identified in the EQA terms of reference, are intended to:

- ❑ Identify key weaknesses in UNFPA evaluation practice, so that strategies can be developed for improvement.
- ❑ Identify useful and concrete examples of good evaluation practices, to guide future evaluation efforts. Examples of good practices could also be integrated into future PM&E training, and other evaluation capacity building events.
- ❑ Serve as a baseline against which the future quality of UNFPA evaluations can be measured.

1.3 METHODOLOGY

Development of the Evaluation Quality Assessment (EQA) methodology was an intensive exercise, including *inter alia* a review of 11 previous meta-evaluations. This Section provides an overview of the method used - further information can be found in the Annexes noted below.

To capture both the quality of UNFPA's evaluation practice, and understand the reasons behind it, a mixed method approach using quantitative and qualitative techniques was adopted. Mixed method approaches are considered to be the best meta-evaluation practices, as they encourage attention to both quality and process, and promote triangulation, and thus credibility, of findings. Lessons gleaned from recent meta-evaluations and academic literature underscore the fact that an exclusive focus on assessing evaluation quality through written reports using rating systems misses important aspects, in particular those related to the use of evaluations. The meta-evaluation also focused on exploring good practices, making them even more useful for UNFPA staff and counterparts. In keeping with this focus, a number of good practice boxes are integrated throughout this report.

1.3.1 Quantitative assessment

The quantitative review focused on assessing evaluation quality through analysis of a representative sample of 60 post-2001 evaluations. Details on sample selection can be found in Annex 2. The meta-evaluation terms of reference required an analysis of the quality of evaluation at UNFPA in relation to three main reference points:

- ❑ **The 1997 Policies and Procedures Manual**, which established policy requirements for UNFPA that were in force for the period under review. Nine quality criteria address these requirements.
- ❑ International good practice, drawing on evaluation standards such as those of the Joint Committee on Standards for Educational Evaluation, and similar exercises, in particular UNICEF (2004). These good practice standards included key evaluative criteria, including the five OECD-DAC evaluation criteria, cross-cutting themes such as gender equality, and conclusions, lessons learned and recommendations. Twenty two quality criteria address international good practices and include the nine criteria mentioned above from the 1997 *Policies and Procedures Manual*.
- ❑ **Standards which address the current programming priorities of UNFPA**, as set out in its 2002 *Strategic Direction Positioning Statement*. Six quality criteria address these priorities.³

The purpose of using three reference points was to analyse evaluation performance against past, current and potential future standards.

3. There is some overlap between the three categories, i.e. some standards fall into more than one category.

Combining the standards from these three reference points, the sample of 60 evaluations was assessed against a total of 28 quality standards. In addition, six quality criteria were developed for assessing evaluation terms of reference, and a further five criteria for assessing the completeness of reports. These standards were compiled into a matrix known as the EQA matrix. Evaluations were rated on individual criteria on a five-point scale. For details of the development of all criteria, a copy of the EQA matrix, details of the rating system, the process of assessing evaluation report quality, and data analysis, see Annexes 2.

One point to bear in mind while reading this report is that because of purposive sampling higher quality evaluations were selected as part of the samples. If random sampling had taken place, rating results likely would have been lower.

1.3.2 Qualitative assessment

The qualitative assessment focused on UNFPA staff views and perceptions of evaluation quality, and processes and factors leading to unsatisfactory or good evaluation practices. It consisted of two elements:

- Interviews with 24 staff at UNFPA HQ, and 14 staff at three CSTs in Addis Ababa, Bangkok, and Mexico City; and
- Six country good practice case studies, and one TSD good practice case study. These involved further interviews with four UN Resident Coordinators, 39 UNFPA staff, 52 counterparts from government and civil society, and interviews with 18 evaluators and 16 beneficiaries. The case study countries were Ghana, Lebanon, Mozambique, Nicaragua, Sri Lanka, and Viet Nam; these were supplemented by visits to India and Iran. Country case studies were selected through a review of the 60 evaluations in the sample, and to ensure geographical representation. In four of the six country cases the good practice study was carried out by two meta-evaluators.

Interviews were also carried out with six staff in the evaluation departments at the HQ offices of IDRC, UNDP and UNICEF. In total 173 people were interviewed. The spread of interviewees across HQ, CSTs, Country Offices, and the range of interviewees across UNFPA staff, counterparts and evaluators, was considered a representative sample on which to base conclusions, and is equivalent to similar exercises. The main interview techniques were semi-structured interviews, either individually with senior staff, or in focus groups. The interview guides are included as Annex 4, and a list of people interviewed in Annex 5. The interview guides were organised around the main questions in the terms of reference; namely, the quality of evaluations reports, the quality of evaluation management, follow-up to evaluations, and quality of support and resources provided by UNFPA.

Semi-structured interviews have been standard practice in development evaluation and the general evaluation field for at least ten years (Kanbur 2003, Bamberger 2000). It is rare to find an evaluation either within or outside UNFPA which does not include semi-structured interview techniques as a central feature, and these have also been noted as keys in guidance for meta-evaluations (Stufflebeam 2001, 2001a). In this report we have included details on the numbers of interviewees making specific comments, in order to substantiate our findings. Where there were clear differences of opinion on key subjects, we have attempted to point this out, and provide a range of views.

The semi-structured interviews included use of Appreciative Inquiry (AI) techniques as a central feature (Cooperrider and Whitney 2002; Hammond 1998). The main focus of AI interviews was on the staff's most positive experiences with evaluation in order to better understand the factors which led to success (see Box 2). Overall interaction at HQ and at the country level provided significant added value to the meta-evaluation, especially in terms of understanding facilitating factors that lead to good quality evaluations, staff attitudes towards evaluation and their needs for technical support.

BOX 2: WHAT APPRECIATIVE INQUIRY IS AND WHY WE USED IT.

Appreciative Inquiry (AI) is an evaluative approach originally used in the private sector, but now increasingly being employed in public sector evaluations as one of a number of participatory evaluation approaches. AI seeks to discover what works well and to understand the elements of success so that they can be replicated.

Appreciative questions that were asked during the UNFPA meta-evaluation included:

- **Think of the time when you first heard or read the evaluation of this project or programme. What made the evaluation so special?**
- **What do you think would make an effective evaluation?**
- **If you could have three wishes for UNFPA for ensuring more of these successful evaluations, what would those wishes be?**

An additional value of AI in this meta-evaluation was that it put people in a positive and open frame of mind. It also captured the knowledge and experience of staff and allowed the identification of good practices.

The purpose of the meta-evaluation and its intended uses were explained in all cases, respondents were ensured of confidentiality, and all respondents were interviewed in a formal setting.

Assessments of the evaluations were staged with country visits so that there was iteration between the two processes. Triangulation between quantitative and qualitative results was carried out mainly to compare and cross-reference findings on evaluation quality, in particular between HQ staff, the Reference Group and the rating against the EQA matrix. It was also useful in relation to identifying both constraining and facilitating factors that can inhibit evaluation quality or contribute to it. Another form of triangulation compared meta-evaluations with equivalent methodologies (see Section 2).

The meta-evaluation adopted a utilization-focused approach (Ginsburg and Rhett 2003; Henry 2003; Patton 1997), concentrating on the needs of intended users. The evaluation team was familiar with the strengths and weaknesses of this evaluation approach, both from their own experience and from academic literature. Despite the trade-offs involved with a utilization focus, the evaluation team viewed it as the best method currently available to meet the meta-evaluation objectives. This was because the objectives focus on improving organizational learning, highlighting effective practice, and improving quality, all of which require substantial buy-in from UNFPA staff, which utilization approaches can facilitate. Part of the utilization-focused approach was to establish at the start of the exercise a UNFPA Reference Group to guide the evaluation. The meta-evaluation process involved interaction with the Reference Group at key points. The Group was asked to: Approve the Methodological Approach Paper that structured the exercise; validate draft findings and recommendations; and comment at various points during production of this report. Recommendations from interviewees were also sought throughout the meta-evaluation.

Further methodological details can be found in the Annexes:

- ❑ For sampling procedures, development of the EQA matrix, the rating system used, and the process of assessment, as well as a copy of the matrix itself, see Annex 2.
- ❑ Annex 3 includes the sample of evaluations assessed.
- ❑ Annex 4 includes the interview guides used during HQ and country visits.
- ❑ Annex 5 is a list of people interviewed for the EQA.
- ❑ Background to team members can be found in Annex 6.
- ❑ Bias and limitations are contained in Annex 7.

1.4 Organization of the report

The report is organized according to the following main headings:

- ❑ **Section 2** covers evaluation quality as measured against the three reference points in 1.3.1 above, based on the results of rating against the EQA matrix. The following individual criteria are analysed in more-depth: completeness of reports; terms of reference; use and users; good practice in methodology, gender mainstreaming; the OECD-DAC criteria; recommendations and their follow-up; and lessons learned. This Section also analyses whether UNFPA evaluations are asking the 'right' questions with regards to the 2002 Strategic Direction Positioning Statement.
- ❑ **Section 3**, based mainly on country visits and HQ interviews, analyses the evaluation processes. In particular it: Examined the evaluation approaches used at UNFPA; carried out a cost benefit analysis which UNFPA evaluators used and the time allocated to evaluations; and included the amount of support given to national capacity building.
- ❑ **Section 4** covers the quality of evaluation resources developed and supported by HQ.
- ❑ **Sections 5 and 7** are conclusions, recommendations and management response.
- ❑ **Sections 6** covers lessons learned, focusing on factors leading to good practice.

2. Main findings from the assessment of evaluations

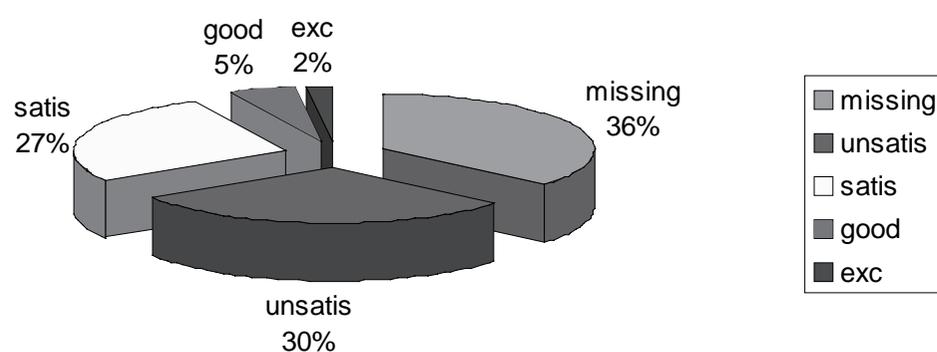
This Section identifies overall strengths and weaknesses in the evaluation sample. It begins with an analysis of rating results against the EQA matrix for the evaluation sample as a whole, and then against two of the reference points noted in Section 1.3.1. The Section then examines performance against individual standards and against the third reference point noted in Section 1.3.1.

2.1 OVERALL QUALITY OF EVALUATION REPORTS

In total, 34 per cent of evaluations rated satisfactory or better across all rating criteria (Figure 2.1).⁴ This is equivalent to similar exercises, including ALNAP's humanitarian action meta-evaluation and UNICEF's meta-evaluation.⁵ In short, our conclusion is that evaluation quality at UNFPA is currently unsatisfactory and requires improvement.

A breakdown by region showed little variation in rating across the five management units, suggesting that evaluation quality is fairly constant across UNFPA.

Figure 2.1 Summary of ratings for 60 evaluations on all criteria



4. All figures in this report were subject to rounding. The high percentage of 'missing' ratings in Figure 2.1 is partly a consequence of three criteria in the EQA matrix being rated as almost totally missing. A missing rating was given when there was no attention to the criteria in the report. An unsatisfactory rating was given when there was attention to the criteria in the report, but this did not meet the criteria requirements as defined in the EQA matrix. For further details on how the rating categories were defined, go to Annex 2. There are also a high number of 'missing' ratings applied to the terms of reference criteria – as can be seen in Table 2.1, because 29 of the 60 evaluations in the sample did not include terms of reference. If the six terms of reference criteria are excluded, the percentage of 'missing' criteria declines to 30 per cent, and the percentage of criteria rated 'satisfactory' or better increases to 37 per cent from 34 per cent.

5. Two main comparators, which used similar methodologies to the UNFPA meta-evaluation, are referred to throughout this Section. These are UNICEF's (2004) meta-evaluation, which covered 75 evaluations; and ALNAP's (2004) meta-evaluation of evaluations of humanitarian action, which covered 197 evaluations over a four-year period. The method of rating used by UNICEF is not entirely clear, but appears comparable to that used in the EQA matrix.

Table 2.1 provides a breakdown of rating by individual criteria, for ease of reference.

RATING OF 60 EVALUATION REPORTS BY CRITERIA (PERCENTAGES)					
Criteria	Assessment				
	Missing	Unsatisfactory	Satisfactory	Good	Excellent
TOR*					
1.a. Purpose/Rationale of the evaluation is clearly defined.	64	4	28	5	0
1.b. Use and users clearly defined.	74	14	8	2	3
1.c. Scope and Focus is clearly identified	52	7	24	16	2
1.d. Methodology specified	63	25	11	2	0
1.e. Roles and responsibilities of evaluators defined	78	12	4	6	0
1.f. Clear Outputs, deadlines, formats specified	70	21	7	2	0
THE EVALUATION REPORT					
3. Quality of Executive Summary.	45	17	27	7	4
4. Purpose of the evaluation is outlined	35	19	38	7	1
5. Use and users clearly defined	58	30	8	4	0
6. Objectives of evaluation clearly defined	13	11	69	6	1
7. Quality of methodology					
7.a. Link to evaluation objectives/questions	10	17	68	4	1
7.b. Good practice	9	64	18	6	3
7.c. Bias and limitations	49	20	23	4	4
8. Consideration given to propriety and ethics	89	8	2	0	1
9. Evaluator values and bias outlined	91	7	2	0	0
10. The project/programme to be evaluated was clearly described	0	65	17	10	8
11. The role and contributions of UNFPA and other stakeholders were clearly defined	10	46	36	3	5
12. Extent of Stakeholders/beneficiaries involvement	82	10	5	3	0
13. Quality of assessment of gender mainstreaming	18	45	21	7	9
14. Quality of assessment of capacity development	5	37	51	6	2
15. Quality of assessment of human rights	67	24	9	1	0
16. Evaluation enables UNFPA to engage in evidence based policy dialogue	35	36	24	3	2
17. Quality of assessment of cultural issues	40	32	21	5	2
18. Quality of assessment of national ownership	5	45	42	7	1
19. Quality of assessment of partnership and alliance building	12	74	7	3	4
20. Quality of assessment of RBM	13	47	22	12	6
21. Quality of assessment with regards to relevance	3	34	54	6	3
22. Quality of assessment with regards to effectiveness	0	51	36	8	5
23. Quality of assessment with regards to efficiency	58	34	7	1	0
24. Quality of assessment with regards to sustainability	20	32	39	8	1
25. Quality of assessment with regards to impact	21	38	33	4	4
26. Quality of conclusions	8	36	47	7	2
27. Quality of recommendations	2	20	71	7	0
28. Quality of lessons learned [^]	-	35	52	11	2

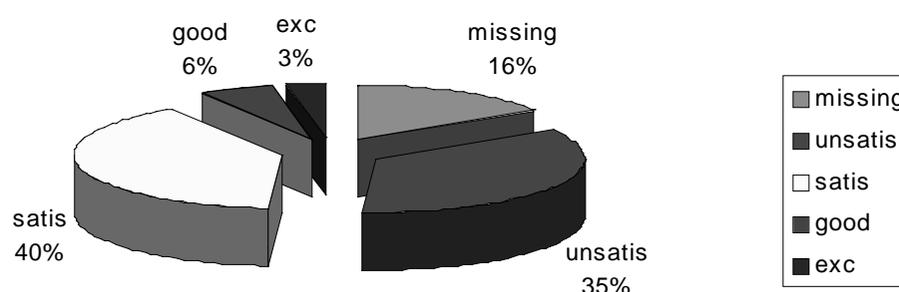
*Based on 28 evaluations which included terms of reference.

[^]Based on 33 evaluations which included a section on lessons learned. 'Missing' was not included for this criteria.

2.1.1 Quality of evaluations as assessed against the 1997 Policies and Procedures Manual

How well do UNFPA evaluations meet the requirements of the UNFPA policy that was in place at the time of their preparation? As noted, one reference point for assessment of quality was the 1997 *Policies and Procedures Manual*. Nine criteria were specifically mentioned in that Manual: gender mainstreaming, capacity development, relevance, effectiveness, efficiency, sustainability, impact, recommendations, and lessons learned. Analysis of these nine criteria illustrates that evaluation performance is satisfactory or better in 49 per cent of cases (Figure 2.2). This is a better performance than for all criteria (Figure 2.1), and for 22 international good practice criteria (Section 2.2.2). Despite this, one in two UNFPA evaluations did not meet policy requirements.

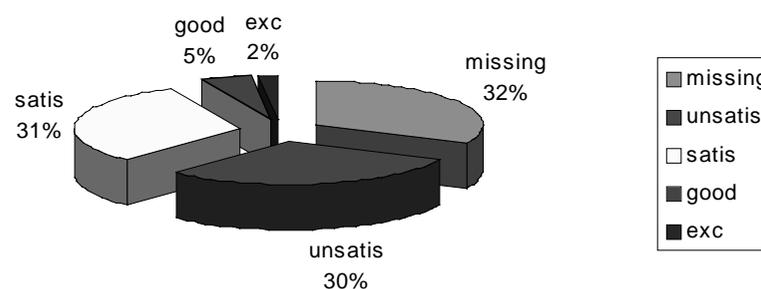
Figure 2.2 Rating by 9 1997 criteria



2.1.2 Quality of evaluations as assessed against international criteria

A further reference point for evaluation quality is accepted international good practice. Twenty two criteria were included in the matrix on this basis (see Figure 2.3 for an analysis of performance against these criteria). The number of criteria rated 'satisfactory' or better on these criteria is 38 per cent, as opposed to 34 per cent for the sample as a whole. Much of this difference can be accounted for by the sub-set of international criteria not including rating of terms of reference.

Figure 2.3 Rating by 22 international criteria



2.2 QUALITY OF ATTENTION TO INDIVIDUAL EQA CRITERIA

2.2.1 Quality of terms of reference

Quality of terms of reference was assessed according to the following six criteria:

- Purpose/rationale of the evaluation is clearly defined;
- Use and users are clearly defined;
- Scope and focus is clearly identified;
- Methodology adequately specified;
- Roles and responsibilities of evaluators defined; and
- Clear outputs, deadlines, and formats specified.

Fifty two per cent of the sample, or 31 evaluations, included a terms of reference, so the analysis is based on this sub-set. The average combined assessment across the six criteria used in assessing the Terms of Reference is given in Table 2.2.

Table 2.2 Average rating on six terms of reference criteria

Category	Percentage
Excellent	2
Good	11
Satisfactory	28
Unsatisfactory	29
Missing	30

Quality of terms of reference was rated higher than the evaluations, with 41 per cent rated satisfactory or better for the terms of reference as opposed to 34 per cent for the evaluations. Across the six criteria, the terms of reference were rated highest on defining the purpose/rationale of the evaluation (72 per cent rated satisfactory or better); and clearly identifying the scope and focus (82 per cent rated satisfactory or better). Specification of methodology was rated lower (74 per cent rated unsatisfactory or missing); as was clear identification of use and users (73 per cent rated unsatisfactory or missing). Terms of reference were overall unsatisfactory. The fact that 48 per cent of evaluations did not include a terms of reference is an indicator of the lack of recognition of their importance.⁶

Is there a correlation between good quality terms of reference and a good quality evaluation? In order to test this hypothesis, average ratings on the six terms of reference criteria were calculated. For 11 evaluations where the terms of reference received a rating of eight or more, the average rating for the report on criteria 3-28 was 1.26. For the remaining 20 evaluations containing terms of reference, the average rating for the evaluation on criteria 3-28 was 1.14.⁷ It can be concluded that higher quality terms of reference contribute to higher evaluation quality.

6. UNICEF (2004: 19) similarly notes: 'The terms of reference, in addition to being unsystematic and having gaps, were generally skimpy.'

7. Data was analysed as follows. An average score was calculated on criteria 3 to 28 for the sample as a whole, with no weighting allocated to criteria, and missing =0, unsatisfactory=1, satisfactory=2, good=3, excellent=4. A similar average was calculated for terms of reference.

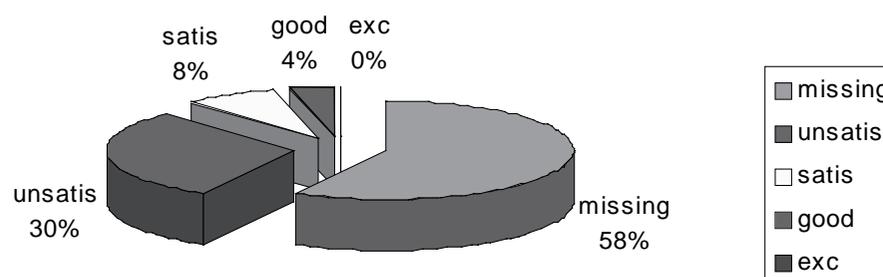
2.2.2 Completeness of evaluations⁸

Fifty two per cent of the evaluations included an Executive Summary, while 48 per cent included a bibliography, and 54 per cent included a list of people met. Data collection instruments were included in 32 per cent of cases. This is equivalent to the UNICEF (2004) meta-evaluation, which found that more than half of its sample did not contain an Executive Summary, and only around half annexed a terms of reference. Ensuring completeness of evaluations is an 'easy fix' as it requires limited extra effort on the part of evaluators, and limited extra screening on the part of Country Offices.

2.2.3 Identification of use and users

As can be seen in Figure 2.4, lack of clear information on intended use and users was one of the weakest areas in the evaluations, with 88 per cent of them rated missing or unsatisfactory. Defining intended use and identifying potential users is central to determining how the evaluation findings and recommendations can be effectively utilized. This shortcoming is a common issue with evaluations, and ALNAP's (2004) review of evaluations of humanitarian action had similar findings over a four-year period.

Figure 2.4 Clear definition of use and users



2.2.4 Quality of methodologies employed

Evaluations were assessed across two main areas:

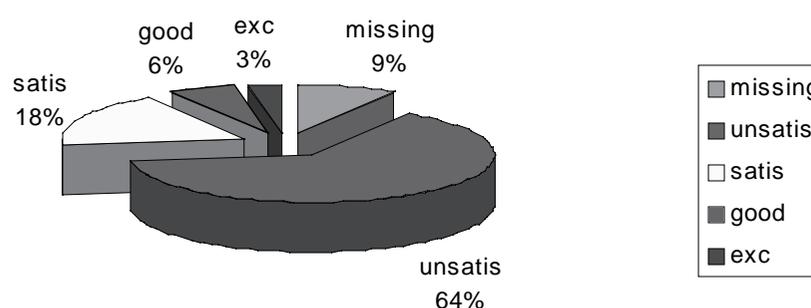
- ❑ whether the methodology followed good practice (i.e. detailing of methods for triangulation, inclusion of adequate description of data collection methods and analysis, inclusion of adequate description of sampling, explicit attention to gender and marginalized groups); and
- ❑ quality of the description of bias and limitations.

Results for good methodological practice can be seen in Figure 2.5, with 27 per cent of evaluations rated satisfactory or better, and 64 per cent unsatisfactory. Nine per cent of the sample included no methodological details.

If measured against good practice, UNFPA evaluations are unsatisfactory. However, neither the UNICEF nor ALNAP meta-evaluations measured up in terms of good practice. The UNICEF report noted (2004: 20): "UNICEF evaluations varied in their methodologies but overall were impressionistic rather than rigorous. The most common methodology was unstructured discussions between evaluators and persons delivering the program, during short site visits... Contact with the treatment (target) group was frequently brief and superficial."

8. The remainder of this Section covers the total sample of 60 evaluation report.

Figure 2.5 Good practice in methodology



As might be expected, evaluations with higher rated method sections were also higher rated for the evaluation as a whole. The 13 evaluations which rated satisfactory or better on good practice in methodology scored an average of 1.24 for the report as a whole, as opposed to 1.11 for the remaining 47 evaluations.⁹ It is clear that greater attention to methodology will lead to a stronger report overall.

The fact that evaluations tend to be superficial in their descriptions of methods used undermines the credibility of UNFPA evaluations. In some cases this could be remedied by better description, for instance in cases where the evaluation used a rigorous method but failed to adequately describe it. In other cases evaluators and evaluation managers do not appear to have a good sense of what constitutes good evaluation practice.

Two particular methodological gaps were noted: Lack of attention to gender equality, and marginalized populations. If there is no specific mention of these areas in evaluation terms of reference, then it is unlikely they will be included in evaluation methodology and practice. Issues of gender equality and marginalized populations were referred to in only 25 per cent of evaluations which included methodologies. Underscoring this shortcoming, UNFPA evaluators rarely canvassed the opinions of poorer groups, and tended to be non-participatory. Eighty two per cent of evaluations were rated as 'missing' against the criteria: "extent of stakeholders/ beneficiaries involvement in the evaluation".

Attention to gender issues in the method section does lead to greater attention to gender in the evaluation report. Of the 13 evaluations which specifically address gender in their method section, 54 per cent were rated satisfactory or better on the criteria "quality of assessment of gender mainstreaming", as opposed to 33 per cent for the remaining 47 evaluations.

2.2.5 Is UNFPA evaluation relevant to its Strategic Framework?

This Section answers the questions in the EQA terms of reference: "In light of current organizational goals and priorities, are UNFPA evaluations presently evaluating the 'right' kinds of things in the 'right' kinds of ways?" In order to answer this question we included six specific criteria in the EQA matrix. This focus is related directly to current organizational priorities as stated in the 2002 Strategic Direction Positioning Statement.

9. See footnote 7 for methods of calculation.

Across the six criteria, assessment of capacity development was rated highest, with 59 per cent satisfactory or better, and assessment of partnership and alliance building lowest, with 14 per cent satisfactory or better. Attention to cultural issues with 28 per cent rating satisfactory or better, was also below the average for the sample as a whole. However, there were two good practice examples identified in relation to cultural issues, one from Sri Lanka, *IEC and Advocacy in Support of RH - A Rapid Assessment*, and the other from Panama, *the Evaluación de Proyecto PAN/00/P01 "Fortalecimiento de la Familia y Mejoramiento de la Salud Reproductiva del Pueblo Ngöbe"*. The latter was assessed as good practice because it: highlighted the conceptual framework of 'interculturality' as the evaluation's point of reference; illustrated that the baseline survey inadequately explored socio-cultural and gender issues and that this impacted on the relevance of project design; explained how the project was designed based on value judgments regarding gender rather than on evidence and how that led to an exaggerated focus on certain preconceived gender issues which ultimately could create conflicts in the community; and analysed how the project design process did not draw on positive cultural assets in the target population.

Table 2.3 Rating of 60 evaluation reports by UNFPA specific criteria (percentages)

Criteria	Assessment				
	Missing	Unsatis- factory	Satis- factory	Good	Excellent
Quality of assessment of capacity development	5	37	51	6	2
Evaluation enables UNFPA to engage in evidence based policy dialogue	35	36	24	3	2
Quality of assessment of cultural issues	40	32	21	5	2
Quality of assessment of national ownership	5	45	42	7	1
Quality of assessment of partnership and alliance building	12	74	7	3	4
Quality of assessment of RBM	13	47	22	12	6

Figure 2.6 Rating against 6 UNFPA specific criteria

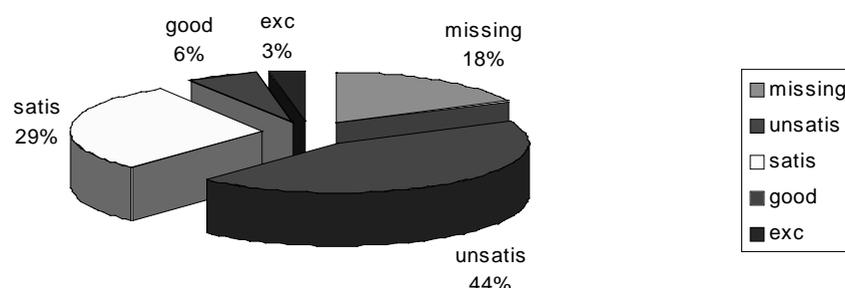


Figure 2.6 provides aggregate percentages for the six specific criteria. It is noteworthy that only 18 per cent were rated as missing, meaning that a large majority of reports addressed these issues, even if 44 per cent did not address them satisfactorily.

Three reports were rated as excellent in terms of quality of assessment as regards results based management (RBM):

- ❑ From Ghana: Strengthening the Implementation of an Innovative and Culturally Sensitive Integrated Community Based Reproductive Health Service Delivery in the Upper East Region of Ghana.
- ❑ From TSD: Evaluation of the Averting Maternal Death and Disability Programme.
- ❑ From Myanmar: Programme Assessment and Review, UNFPA Special Programme of Assistance to Myanmar (2002-2005).¹⁰

One of the common themes of these three evaluations is that they all used international evaluators, who might be expected to be more familiar with RBM concepts. These evaluations stand out as excellent because they analysed the use of indicators, and examined the existence of results-oriented PM&E systems, and whether information produced by these fed into decision-making.

As can be seen in Table 2.3, 71 per cent of evaluations were assessed as unsatisfactory or missing on the criteria of "Evaluation enables UNFPA to engage in evidence based policy dialogue". From a design perspective, it was rare for terms of reference to specify evaluation use. And there is no evidence to suggest that UNFPA evaluations are being designed to feed into policy dialogue.

However, from some of the country case studies evaluations were used for generating policy dialogue. But these were included as part of overall policy discussions, not a reflection on the evaluation process. For example, in Lebanon, Sri Lanka, and Viet Nam UNFPA Country Office staff described a process of policy dialogue taking place over a number of years – ten years in one case – into which the evaluations under review comprised one component. Where evaluations involved a participatory process between UNFPA and the host government, policy dialogue around the evaluation findings and recommendations was inevitable. One example of policy change brought about directly as a result of an evaluation was in Ghana (this is highlighted in Box 3).

In relation to evaluating policy dialogue and advocacy (as opposed to evaluations being used for policy dialogue), given the current focus of evaluations at the output and project level, there is limited information on the results of policy dialogue and advocacy. Our discussions with HQ and Country Office staff -- in particular UNFPA Resident Representatives and Assistant Resident Representatives – made it clear that while they are continually involved in policy dialogue, the process and results are not systematically tracked. Given the sensitivity of many areas included in policy dialogue – such as human rights and gender concerns– it is not surprising that there is limited formal tracking. Currently, there is limited experience in the evaluation of policy dialogue in the wider field of evaluation.

10. One of the authors of this evaluation was also interviewed at the CST, Bangkok.

● **BOX 3: HAVING A SEAT AT THE TABLE: LEVERAGING EVALUATION WITHIN POLICY DIALOGUE**

UNFPA is participating in Ghana's health sector SWAp, including a performance based monitoring and evaluation system, which routinely collects data on a variety of health indicators. The PM&E system also commissions independent evaluations, which serve to validate the accuracy and reliability of sector performance data; and includes provisions for policy-related thematic evaluations on an annual basis. Every six months a summit is held to review results and discuss emerging issues.

Good quality data provided through the Reproductive Health Unit of the Ghana Health Service indicated that national safe motherhood indicators had begun to plateau. However, Ghana's Poverty Reduction Strategy did not include a systematic method for addressing reproductive health problems. UNFPA advocated forcefully for a sector wide response to maternal mortality, and subsequently a thematic evaluation was commissioned, involving two CST Advisors from Addis Ababa. Costs were shared within the SWAp arrangement.

One senior Government official commented: "Although the decision to undertake this [thematic] evaluation came through UNFPA's advocacy efforts, it was important that the resulting effort not be seen just as a UNFPA evaluation but rather as a sector-wide effort. This helped ensure broad ownership of the results. Since then, a lot of good things have been happening. There is now a new draft national strategy to address reproductive health that came out of the thematic evaluation, and we have begun to change the way we train our community-based health workers. There is now an opportunity to bring reproductive health issues into the broader development framework."

However, 22 of the 25 respondents¹¹ questioned about this noted that as UNFPA moves deeper into SWAps and general budget support, evaluations will need to include greater attention to the results of policy dialogue. Box 4 illustrates some of the challenges UNFPA faces in evaluation policy dialogue based on discussions with Country Office staff and counterparts in Sri Lanka.

● **BOX 4: EVALUATING POLICY DIALOGUE?**

As part of its Gender Based Violence programming, the Sri Lanka Country Office has been working with a well-respected consultant obstetrician at a Colombo hospital. On the initiative of UNFPA staff, this partner was invited to make a presentation to a Parliamentary Committee reviewing a Bill on Domestic Violence. His presentation on issues of domestic violence apparently had a significant impact on committee members. The UN Gender Working Group (which is currently chaired by UNFPA) subsequently supported the National Committee on Women to lobby for this Bill with key women Members of Parliament (MPs). The Bill included dissemination of research findings on gender based violence, the results of gender based violence on families and screening for domestic violence at health centres. The Bill was passed by parliament in August 2005.

In order to analyse impact, evaluators would need to examine whether the Bill would have passed without UNFPA interventions, including any changes that had been made in the Bill as a result of inputs from UNFPA partners; for example, through interviews with parliamentarians.

11. 14 HQ staff and 11 UNFPA Resident and Assistant Representatives.

In the context of policy dialogue, SWAps and general budget support, attribution was an issue raised during Reference Group meetings, and in Country Office visits. During these discussions two levels for evaluation were considered:

- ❑ **UNFPA's decision-making process**, that is why UNFPA decided to become involved in SWAps or general budget support, and the effectiveness of the route it chose. UNFPA can be held directly accountable for this decision.
- ❑ **The results of the SWAp**. This would be best evaluated jointly by all partners. There will probably be no direct attribution of results to UNFPA, but there can be a plausible assumption that UNFPA's involvement has led to a particular result.¹²

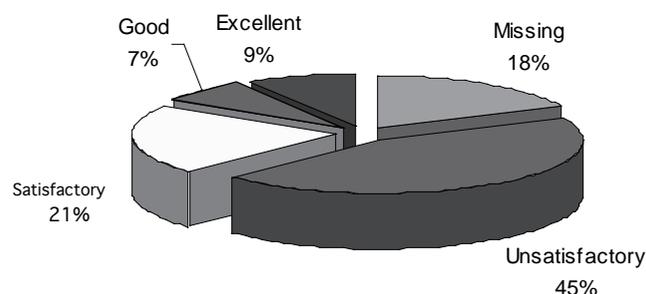
2.2.6 Gender mainstreaming

Assessment of gender mainstreaming involved reviewing evaluation reports for the extent to which:

- ❑ the intervention effectively used sex-disaggregated data; and
- ❑ the intervention promoted gender mainstreaming and improved gender relations.

Both of these areas were unsatisfactory, and the overall rating for this criteria can be found in Figure 2.7.

Figure 2.7 Quality of assessment of gender mainstreaming



It is a troubling finding that 63 per cent of evaluations were rated as unsatisfactory or missing on this criteria, especially given the fact that one of the MYFF 2004-2007 main goals is the promotion of gender equality and the empowerment of women. What this means is that UNFPA has limited evaluative information on gender mainstreaming to support improved programming.¹³ Figure 2.7 also notes that nine per cent of evaluations were assessed as excellent, demonstrating that UNFPA can achieve good practice in this area (two examples of good practice are highlighted from the case study visits in Box 5).

12. This typology was suggested by a senior UNFPA staff member.

13. Other meta-evaluations had similar findings. ALNAP (2003) found that of 127 evaluation reports 26 rated satisfactory or better in terms of attention to gender equality. The UNICEF meta-evaluation found (2004: 33): 'Of all the areas of children's rights addressed in the evaluations we reviewed, gender analysis was weakest, with the exception of political analysis that was indeed rare.'

BOX 5: GOOD PRACTICES IN PROMOTING GENDER EQUALITY: THE CASES OF SRI LANKA AND MOZAMBIQUE

Promoting gender equality is a key element of the Sri Lanka Country Programme, and there is strong commitment to this issue from Country Office (CO) staff. The Country Office has been able to hire evaluators who have sensitivity to, and a strong background in, the issue. Gender is also regularly included as an area of enquiry in evaluation terms of reference. These are the primary factors which have led to good attention to gender issues throughout the CO evaluations, allowing the CO to strengthen its programmatic support to gender.

One of the stated objectives of the evaluation of the Geração Biz programme in Mozambique was "to assess the extent to which gender issues have been incorporated into the programme approach and implementation." The evaluation pointed to the difficulty of recruiting and retaining female peer educators. Obligations at home and relocations arising from marriage were the main causes. Stereotyping by sex was also a problem. The evaluation appropriately made recommendations for the development of new strategies for increased female recruitment, retention and leadership roles. During the EQA case study, a group of ten peer educators who had been involved in the evaluation were asked what the most important results and follow up were to the evaluation. Both female and male peer educators cited the evaluation's positive impact on building momentum for gender equality issues. Some argued that more space had been created for encouraging female participation, others that new advocacy training efforts were more strongly focused on bringing in new girl peer educators, and that extra efforts were being made to appoint females to more senior roles.

2.2.7 Human rights

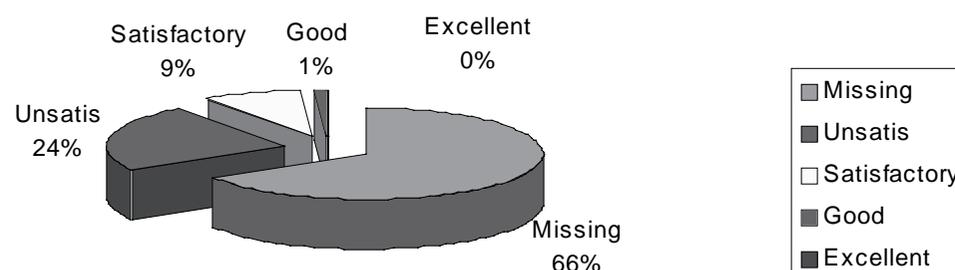
Evaluations were assessed against the extent to which they:

- were informed by the ICPD Programme of Action as listed above, and CEDAW;
- identified the human rights claims and obligations relevant to the project/programme;
- identified gaps in the capacity of rights-holders to claim their rights, and of duty-bearers to fulfil their obligations, including (but not solely) an analysis of gender and marginalized and vulnerable groups;
- addressed identified gaps; and
- monitored results.

In order to achieve a satisfactory rating, evaluations had to meet the first three areas listed above (but not in-depth). As can be seen from Figure 2.8, only 10 per cent of evaluations achieved a satisfactory or better rating, and there was no good practice identified. In 66 per cent of cases there was no reference to human rights. Although attention to human rights was not included in evaluation requirements in UNFPA's 1997 Policies and Procedures Manual, it now provides an overarching framework for UN interventions (including UNFPA) and thus needs greater attention from evaluations.¹⁴

14. See Policy Note on Implementing a Human Rights-Based Approach in Programming in UNFPA. 22 January 2004.

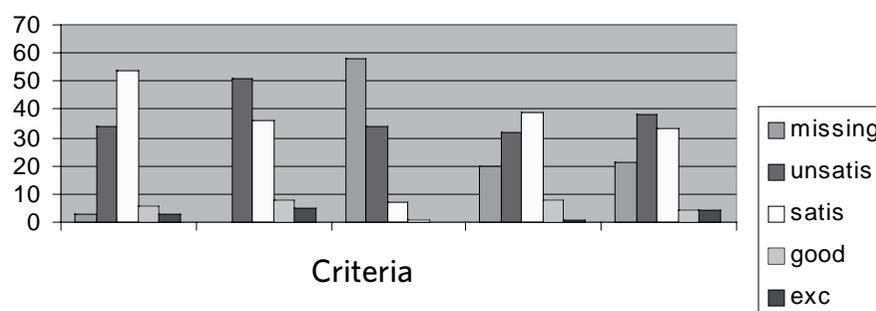
Figure 2.8 Quality of assessment of human rights



2.2.8 Quality of use of the OECD-DAC evaluation criteria

Use of the five OECD-DAC criteria was one of the strengths of the sample, in particular its use of the criteria "relevance" and "effectiveness". Overall, 57 per cent of evaluations were rated as satisfactory or higher on two or more of the DAC criteria. This is consistent with other meta-evaluations, for example the ALNAP review of evaluations of humanitarian action. Comparative ratings can be seen in Figure 2.9.

Figure 2.9 Quality of use of OECD-DAC criteria



Use of 'relevance' was among the highest rating criteria on the EQA matrix, with 63 per cent of evaluations rating satisfactory or better. On the other hand, use of 'efficiency' was comparatively weak, with 58 per cent of reports making no mention of this criteria, and a further 34 per cent rated as unsatisfactory. This is also consistent with findings from both the ALNAP and UNICEF evaluations. The latter found that (2004: 36): "Virtually no [evaluation] compares costs and results. Ninety per cent of evaluations received a zero or one on the scale 0-4 in regard to 'Costs and results should be directly compared to as full an extent as possible.' "

Regarding effectiveness, 49 per cent of the sample was rated as satisfactory or better. Evaluations did not fare well on two of the effectiveness sub-criteria: assessment of coverage, and inclusion of a causality analysis. Lack of attention to coverage, in particular of poorer groups, was a major gap. There was limited information in the evaluations on which sections of the population were being reached by UNFPA programme funds. The MYFF 2004-2007 notes (UNFPA 2004: 10):

“The ultimate results of UNFPA support to, for example, service delivery, will have to be measured by its contribution to overall availability, especially for the poor.”

A similar focus on prioritizing reaching poorer groups was found in planning and reporting documents for country case studies, for example the most recent Ghana, Mozambique, Sri Lanka and Lebanon Country Programmes and Country Office Annual Reports (COAR). Given the links between poverty reduction and reproductive health in the MYFF 2004-2007, UNFPA needs to know whether it is accessing the right target groups. Evaluations should be one key source for this information, through their assessment of effectiveness.

Because evaluations tended to focus at the output level, impact was not well covered, with 41 per cent of evaluations rated as satisfactory or better. UNICEF evaluations received an equivalent rating (2004: 38, 9).

In the TSD sponsored *Evaluation of the Averting Maternal Death and Disability Programme*, an evaluation of a grant from the Gates Foundation to Columbia University for a programme entitled ‘Averting Maternal Death and Disability’, and implemented by UNFPA and five partners, was considered good practice in use of the DAC criteria. As well as good attention to effectiveness and relevance, the evaluation was also rated good on efficiency and impact. In relation to the former it found (p. 14): “The 1993 World Bank World Development Report noted that maternal health services are among the five most cost-effective interventions in low-income countries, costing \$60 per DALY and averting three per cent of the disease burden. This compares favourably with other investments such as family planning which costs \$100 per DALY and averts only one per cent of the disease burden...” The evaluation cited excellent attention to impact, including: a section on challenges to measuring programme impact on maternal mortality; policy changes brought about by the programme; and changes in maternal mortality patterns with clear links to the programme.

2.2.9 Quality of lessons learned

Lessons learned were defined as a contribution to general knowledge with implications for future action, and as learning from experience that is applicable to a generic situation, rather than to specific circumstances. Thirty-three of the evaluations included a section on lessons learned. Of these evaluations: 35 per cent were rated unsatisfactory, 52 per cent as satisfactory, 11 per cent as good, and 2 per cent as excellent. This is higher than the average for the sample, suggesting that evaluations are stronger in drawing inferences and broad conclusions from their findings rather than developing methodologies and establishing findings. However, only two evaluations directly involved stakeholders in the identification of lessons learned.

2.2.10 Recommendations and their follow-up

Seventy one per cent of UNFPA evaluations were rated as satisfactory, and seven per cent as good, in terms of quality of recommendations.¹⁵ To be rated satisfactory, evaluations needed to meet the first four sub-criteria, out of a total of six:

- follow logically from the evaluation findings and conclusions;
- be relevant to the programme/project;
- be clearly stated and not broad or vague;
- be realistic and reflect an understanding of the commissioning organization and potential constraints to follow-up;
- be prioritized with a timeframe for follow-up and suggest where responsibility for follow-up should lie; and
- be formulated with participation of key stakeholders.

Only two of the 60 evaluations in the sample met the fifth sub-criteria. Evaluations contained long lists of recommendations, with no priorities and no allocation of responsibility or timeframe for follow-up.

It is not usually possible to determine from reading evaluation reports the extent to which recommendations have been followed up. However, it is likely that given the overall unsatisfactory quality of UNFPA evaluations, there is limited follow-up. This was also the message received from semi-structured interviews. Of those responding in relation to follow-up from evaluations, 41 respondents thought there was inadequate follow-up, while only 21 thought follow-up was adequate. However, 19 of the 21 responses were from evaluators or government staff involved in the good practice case studies, and they were referring specifically to the good practice evaluation. On the other hand no HQ staff thought there was adequate follow-up to evaluation recommendations.

In the good practice case studies there had been extensive follow-up to recommendations, in particular where the evaluators had adopted a participatory approach, such as in Lebanon, Mozambique, and Viet Nam. Country Offices have also been experimenting with follow-up matrices, and the examples of Mozambique and Viet Nam are given below. Increasingly, such matrices are becoming standard in evaluation practice.

Meetings towards the end of the evaluation with key stakeholders were also useful for ensuring follow-up to recommendations. For example, in Mozambique three meetings took place at the end of the evaluation, involving one day-long working session with government counterparts and other implementing partners, to elaborate an operational plan to implement recommendations and draw up an outline for programme expansion. These meetings took place right after the evaluation, while the findings and recommendations were still fresh, and were also followed-up systematically every three weeks with working group sessions involving the Ministries of Health, Education and Youth, NGOs and key donors.

15. For similar findings see UNICEF (2004) and ALNAP (2004).

Recommendation follow-up matrix Mozambique

Recommendation	Incorporation in the Geração Biz Programme according to the external evaluation
Transport should be made available to peer educators in order to cover their large working areas.	We are developing an operation study to evaluate the use of bikes. In Gaza, we had a lot of problems on this issue. To do that, we have to develop a new strategy.
Opportunities should be identified to involve youth in more aspects of the programme.	The recommendation has been incorporated, and can be seen in the strengthening of youth participation in various parts of the programmes results and aims.

Recommendation follow-up matrix - Vietnam

#	Recommendation/findings	Organization	Action and deadline
A1	To form the basis for monitoring and evaluating CP6's implementation in late 2005, the Government and UNFPA should update the OVI by consulting findings from the Baseline survey on current status of RH care service provision and utilisation in various provinces, as well as existing data on current implementation and the potential for future implementation and funding sources (both regular and mobilised funds).	UNFPA & Ministry of Planning	The OVI of 16 outputs will be revised and confirmed before 11/2004 for reallocation of budget, if possible. The LFA of 11 provinces will be updated/revised, if needed before 11/2004, by province.

2.2.11 Evaluation of training

While this was not an area covered in our terms of reference, we have included a Section on evaluation of training because it was a major focus of the evaluations reviewed; 66 per cent of the sample included references to training, and training was the main evaluation focus in at least 25 per cent of the samples.

In the evaluation field there is generally an accepted model for the evaluation of training, known as the Kirkpatrick model (Kirkpatrick 1998). This assesses four potential outcomes:

- Level 1: Reactions. This focuses on what participants think and feel about the training; including their personal appraisal.
- Level 2: Learning. This focuses on what trainees didn't know before – the extent to which their knowledge levels have increased.
- Level 3: Behaviour. The focus here is on the extent to which trainees use their new knowledge in their own work; ways in which job performance has improved.
- Level 4: Results. This focuses on results at the organizational level, and the extent to which training has contributed to meeting organizational goals.

Preskill (2001) notes that most training interventions are evaluated at the reaction and learning levels. The main focus of UNFPA evaluations was at the reaction level, or Level 1. Recording participants' reaction to training through surveys carried out at the end of the training session was the most common form of assessment.

During the Viet Nam case study visit we found that the Country Office conducted monitoring of training activities in 11 Provinces in 2004, which fed into its Mid-Term Review. The objectives of this were to: Identify strengths and weaknesses of the training process and management; explore the trainees' application of knowledge and skills to their work; and provide recommendations for improvement.

Data sources included project documents, training course pre-test and post-test results for 1,650 trainees, supervision reports, and a survey sent to 2,705 trainees by the assessment team, with a 76 per cent response rate. Consequently, monitoring activities covered the first two levels of Kirkpatrick's model, and appeared to have partly covered the third level.

3. Quality of evaluation organisation and processes

This Section draws mainly on country case study visits which, as noted, were selected because they represented good practice among the evaluation sample; and on semi-structured interviews. A profile of the case study evaluations is provided in Table 3.1. Interview guides and a list of people interviewed can be found in Annexes 4 and 5.

Table 3.1: Profile of good practice case studies

Country	Focus of evaluation	Carried out by	Evaluation Cost (US\$)	Length of evaluation
Ghana	Reproductive Health	One international and one national	11,771	14 days
	Reproductive Health	One international and one national	10,415	15 days
Iran*	Country Programme	One national and assistant	10,000	75 days
Lebanon	Reproductive Health	Six nationals	30,000	60 days
Mozambique	Reproductive Health	Two nationals Two internationals	46,800	21 days
Nicaragua	Reproductive health	One national and technical team of four	13,047	40 days
	Rights and Gender Equality	Two nationals	4,800	30 days
Vietnam	Mid-term	Three nationals	17,500	60 days
Sri Lanka	Reproductive Health	Three nationals and one international	23,700	21 days
	IEC and advocacy	Two nationals and one international	39,500	42 days
	Reproductive Health	One international	21,000	30 days
TSD	Gender Based Violence	One international	40,335	56 days

*Iran was not a full case study.

3.1 THE EVALUATION CULTURE AND APPROACHES AT UNFPA

No staff questioned the importance of strengthening the evaluation function at UNFPA. Country Offices visited have been innovative and produced useful evaluations which have reshaped programming to make it more effective. The general support throughout the organization and the ability of UNFPA to achieve good practices across a range of Country Offices suggests there is a base on which to build better evaluation practice.

However, our finding from interviews, in particular at HQ where we focused the discussion on evaluation, was that UNFPA currently lacks a strong evaluation culture, and evaluations do not have a strategic role in decision-making. Evaluation is not perceived by most staff interviewed as having a coherent function at UNFPA, nor are evaluations being utilized for improving programme/project

impact. This suggests the need to improve UNFPA's organizational evaluation strategy. As one staff member told us: "At UNFPA we need a common vision of what evaluation is all about." Ten HQ interviewees agreed with the comments above, as did an additional 17 respondents from the COs. No interviewees expressed contrary views.

Both DOS and country level evaluations were perceived as being too focused on accountability, as well as too threatening, suggesting the need for more emphasis on models that highlight participation and lessons learned. Respondents told us there is a culture of blame, rather than a culture of learning. Of those respondents that were asked about how to improve evaluation quality at UNFPA, 78 out of 108 noted that evaluations should be more participatory, with no contrary views. Respondents said they need better guidance on what kinds of evaluation models are relevant in different contexts, and the benefits and drawbacks of different approaches. CSTs were also uncertain as to their role regarding evaluation.¹⁶ These findings explain some of the reasons for current low scoring against the EQA matrix.

BOX 6: BUILDING NEW PARTNERSHIPS FOR EVALUATION: THE CASE OF VIET NAM

In mid 2004, the UNFPA Country Office and the Government of Viet Nam decided to conduct a joint mid term review (MTR) of the Sixth Country Programme (CP). The resulting evaluation process took place over a two month period. Most respondents noted that it was the participatory process of the MTR that they valued most. In effect, the country office used a utilization focused approach concentrating on ownership of the evaluation process, and results, by key stakeholders. A central feature of this approach was the establishment of a joint Government of Viet Nam/UNFPA Working Group, chaired by a senior official in the Ministry of Planning and Investment. Representatives from key Government ministries and most UNFPA programme staff were included. Working group meetings were held on a weekly and sometimes even bi-weekly basis, each one taking a couple of hours or more as every detail was discussed and debated. A final one-day wrap-up workshop was also held, which was attended by some 70 participants, including regional counterparts.

During the Appreciative Inquiry interviews, the following was noted by respondents:

"Sitting at all those meetings was certainly tedious at times, but walking into the room, one very much had the impression of we Vietnamese sitting and discussing our own issues and problems."

"In the past when we undertook mid term reviews by ourselves as a form of reporting exercise, the reports tended to sit on the shelf. It was very difficult to get any action. Now we get action."

"The whole experience allowed us to develop better relations with one another."

The Government cited the following advantages: An increased sense of the Government's stake in the UNFPA Country Programme; strengthened partnership between UNFPA and the Government on the Country Programme and ICPD follow up; and a greater willingness to be held accountable for taking remedial action on the evaluation's recommendations. Good follow-up to recommendations has been one notable success stemming from the participatory process.

16. All CST interviewees were questioned concerning this, and we received an additional three responses, for a total of 17 respondents. Of these 17 respondents, 11 supported the comments above, with one contrary opinion.

In addition, COARs do not appear to integrate systematically the findings from evaluation reports. An exception from the case study countries is the Sri Lanka Country Office, which demonstrates the ways in which evaluations can be effectively used in strategic planning and in the COAR.¹⁷

UNFPA evaluations for the most part use standard methodologies, relying on document review and interviews with key stakeholders. We found only one evaluation using quasi-experimental design. As noted, few adopt participatory methods, or systematically survey beneficiaries or marginalized groups.

Country Offices have been exploring more participatory and utilization-focused approaches, focusing on learning and counterpart ownership of evaluation processes, findings and recommendations.

● **BOX 7: SEEING OUR FOOTPRINTS IN THE NEW COUNTRY PROGRAMME: A PARTICIPATORY EVALUATION IN IRAN**

In April 2004, UNFPA neared the end of its 3rd Country Programme (CP). With a new CP under preparation, the UNFPA Representative felt that an Operational Evaluation of the Country Programme would complement work already done on the new CP. An Operational Evaluation was planned to look at the way UNFPA does business; how it makes technical assistance available; how it manages and executes projects; and, how it selects its intermediaries.

The UNFPA Country Office wanted stakeholders to participate actively in the evaluation, so they would have a strong sense of buy-in and assume ownership of its results. As the Representative explained: "if you present stakeholders with a ready made report, buy-in is minimal." UNFPA agreed with the Participatory Evaluator/Facilitator to undertake a series of workshops involving 17 key stakeholders: National Project Directors, UNFPA, and some senior national project staff.

Over the course of 2.5 months, five one-half day workshops were held. The first workshop saw the Representative take a leading role in providing background, and putting stakeholders at ease by explaining that this was not an exercise where they were going to be judged. As the Representative commented: "We prepared them for the exercise." We shared the terms of reference, planned the process of evaluation and scheduled it to get them involved." The dual role of the consultant was also explained - she would be an evaluator at times, and at times a facilitator of the workshops. At the first workshop, stakeholders broke into small groups and discussed the type of participation they would like to see. Subsequent workshops looked at: Technical Assistance Arrangements; Implementation and Execution Modalities; and Financial Aspects. The final workshop was used to present findings, conclusions, lessons learned, and recommendations.

As for impact, the Evaluator exclaimed: "We can see our footprints in the new Country Programme." The Representative added that: "We have established a roster of local consultants and NGOs to serve as a primary source of technical assistance for the country programme implementation in place of CSTs whose focus has shifted to strategic interventions (CCA, UNDAF, MDGs, SWAp, etc.). We are planning to team up CST advisors with local consultants to build the latter's capacity. We are also recruiting national project officers who will integrate the monitoring and evaluation function into their responsibilities."

17. COARs reviewed were from Lebanon, Vietnam, Mozambique and Ghana.

As one Country Office staff member put it:

“Because of the participatory nature of the approach our partners were more receptive to the findings of the evaluations, even to the criticism. The government gets criticized a lot so it is important to make evaluation a positive experience rather than something threatening.”

Since respondents noted that the evaluation approach adopted is a key issue for UNFPA, we have included two good practice case studies on successful participatory approaches in Boxes 6 and 7, from Viet Nam and Iran.

3.2 Who carries out UNFPA evaluations

Fifty-four evaluations included information on the institutional background of the evaluators. There are three main points to note concerning evaluators:

- ❑ Of the sub-sample of 54 evaluations which included details on evaluators, 46 per cent were carried out solely by national evaluators, a further 30 per cent by national and international evaluators, with the remaining 24 per cent carried out solely by international evaluators, CST advisers, or a combination of these. Given the preponderance of national evaluators, individually or in combination with international evaluators (76 per cent of cases), a key factor for improving the quality of UNFPA evaluations is for Country Offices to access high quality national evaluators. This was reiterated by UNFPA staff in Lebanon, Sri Lanka, and Viet Nam. But there is a catch - UNFPA has to compete for these evaluators with other UN agencies and NGOs. One long-term solution involves increasing national capacity, both the capacity of evaluation functions in counterparts as well as the pool of qualified evaluators.
- ❑ Higher quality evaluations appear to be produced by mixed teams of national and international evaluators. The six evaluations involving mixed teams of national and international evaluators (excluding CST advisers) received a 50 per cent satisfactory or better rating, as opposed to 34 per cent for the sample as a whole. These six evaluations received an average score of 1.3 across criteria 3-28 on the EQA matrix, as opposed to an average score of 1.14 for the remaining 48 evaluations (the basis for calculations is in footnote 7). While this conclusion is based on a small sample, it is in line with findings from the ALNAP meta-evaluation (2004). However, this is not always the case - for example in our Lebanon, Viet Nam and Iran good practice country visits we found evaluations conducted solely by highly skilled national evaluators.
- ❑ CST advisors were involved in 22 per cent of evaluations. CST advisors are attractive evaluators for Country Offices because they know UNFPA, are located in the region, and do not charge for their services except for travel and per diems. However, while CST advisors have strong technical backgrounds, they do not always have the requisite skills to conduct evaluations, have not received recent training on evaluation, and/or are available for very short periods of time.

There does not appear to be a standardized process for selection of evaluators. Evaluation reports say little about selection processes. Our country visits revealed that selection is mainly through networks and word of mouth; not necessarily an issue where specialized evaluations can be carried out and where the potential pool of evaluators is limited. The Lebanon Country Office has developed an innovative tendering process which could be used by UNFPA in general. Not only did this process ensure transparency, but it also promoted ownership by counterparts of the evaluation because of their early involvement in developing tendering guidelines, selecting evaluators, and developing the terms of reference. This is highlighted in Box 8.

BOX 8: SELECTION OF EVALUATORS BY LEBANON COUNTRY OFFICE

In the Lebanon case, the participatory approach began with the tendering process, which was used to build ownership of the evaluation from its inception. The Country Office developed a scoring system to rate evaluation tenders. Intensive discussions were held concerning categories to be included in this scoring system, and the weight to be given to different elements, over the period of about a month, involving UNFPA and three government departments. This focused discussion on the expectations of the various parties as to what they hoped the evaluation would cover, and promoted participation and consensus building. Even before the evaluation had begun, ownership over the evaluation process had already begun to be established. Various counterparts were also involved in writing the evaluation terms of reference.

Seventy per cent of scores were based on technical areas, and 30 per cent on financial aspects. As well as promoting participation of the various partners, the tendering framework was also transparent and allowed the Country Office to respond to one of the tendering parties who enquired about the decision-making process for awarding the contract. This is a case of good practice which could be used by other UNFPA Country Offices.

3.3 Length of time devoted to evaluations

The average length of time devoted to evaluations, based on a sub-sample of 21 evaluations, was 34 days (not including good practice evaluations). This figure is biased by the fact that four of these 21 evaluations collectively took 362 days to complete. If these four evaluations are removed, the average length of time for the remaining 17 was 20 days. Time allocation is dependent on individual Country Offices - some evaluations were just a few days, others required several months. While evaluation quality depends on several factors, there appears to be a correlation between length of time spent on the evaluation and evaluation quality -- the average length of time spent on the 11 evaluations included in the good practice case studies was 42 days. The main limitation that evaluators gave was lack of adequate time. In the 28 evaluations which included a section on 'bias and limitations', lack of time was mentioned as a constraint in 20 evaluations, or 71 per cent of cases. In addition, 17 evaluators out of the sample of 18 interviewed mentioned lack of time as a constraint to producing good quality evaluations.

Time spent on evaluations is too short because it does not allow for:

- Carrying out of basic evaluation functions, such as document review, and interviewing and feedback to key stakeholders.
- An examination of key questions raised in the *Strategic Direction Positioning Statement*.
- A participatory process with counterparts.
- Systematic consultation with beneficiaries.

One of the reasons that a number of evaluations were too short was the involvement of CST advisors, who are usually only available for limited periods of time. A further reason is that some Country Offices were not clear on what constitutes an optimal amount of time for an evaluation.¹⁸

18. Compare UNICEF (2004: 27): 'Many evaluations we reviewed were conducted in a short time. For the most part, this is false economy.'

Length of time dedicated to the evaluation is partly dependent on resource allocation. For our good practice case studies, between one and five per cent of the programme budget was spent on the evaluation (see Table 3.1).¹⁹ Of course spending more time on evaluations does not guarantee better practice, but it is one of several factors that will contribute to this objective. Nor should UNFPA HQ be prescriptive about this issue; rather, evaluation managers should be provided with clear information on what is an adequate length of time for different types of evaluation.

During the Viet Nam country visit we discussed the trade-offs involved in investing in participatory, but time-consuming, evaluation approaches with the UNFPA Country Office, counterparts and evaluators. In this instance, the good practice case was the Mid-Term Review of the Country Programme. In boxes 6 and 7 we have highlighted some of the findings related to the length of time needed for participatory evaluations, and the potential advantages of this approach.

3.4 Cost-benefit analysis

The meta-evaluation made a general assessment of the costs and benefits, adapting the methodology used in the World Bank (2004) review of influential evaluations. In five out of seven good practice case studies savings or likely savings in programming were higher than the cost of evaluation.

In Viet Nam the evaluation led to more efficient use of resources within the Country Programme. The evaluation cost US\$17,500, excluding staff time; however, all parties agreed that the benefits far outweighed the costs. One major benefit was an increase in disbursement rates. Prior to the evaluation, the Country Programme was experiencing average disbursement rates of 65 to 70 per cent, but following the evaluation, the rate increased to nearly 100 per cent, as funds were re-allocated away from poorer performing activities. In Mozambique the evaluation cost was \$46,800, and the evaluation had a direct impact on the future planning of the Geração Biz programme, attracting an additional US\$15 million from various donors, funds that might not have materialized otherwise.

3.5 SUPPORTING NATIONAL CAPACITY

Both the UN Evaluation Group Norms of Evaluation in the UN System and the UNFPA Policy and Programme Manual, as well as donors, note the importance of supporting national capacity. The former comments (UNEG 2005: 2):

“Resolutions of the General Assembly and governing bodies of UN organizations imply particular characteristics for the evaluation function within the United Nations system. Evaluation processes are to be inclusive, involving governments and other stakeholders. ... In addition, the General Assembly has requested that the UN system conducts evaluations in a way that fosters evaluation capacity building in member countries, to the extent that this is possible.”

The latter notes (UNFPA 2004: 2):

“Wherever possible, monitoring and evaluation efforts should build on existing mechanisms and processes within the UNCT and the national context. If effective national results-based monitoring and evaluation systems are already operational or are being fully supported by UNFPA or other do-

19. Very few of our sample of 60 evaluations included the cost of the evaluation so we were unable to calculate average cost.

nors, for instance, in the context of SWAps and PRSPs, UNFPA should use these systems as feasible and thus avoid creating a parallel system. Where such systems are not fully functional or non-existent, UNFPA should seek to work together with other United Nations organizations and partners in creating the necessary capacity through such initiatives as the UNDAF, SWAps or PRSPs to develop such systems.”

The context for supporting national capacity is that the quality of evaluators is one key factor leading to better quality evaluations. Over three quarters (76 per cent) of evaluations in the sample were carried out by national evaluators (or mixed teams including national evaluators). Clearly, improving national capacity will lead to an improvement in UNFPA evaluation quality (see Section 3.2).

There is some evidence, however limited, in the case study countries of attempts to support national evaluation capacity. In Lebanon, training for counterparts on monitoring and evaluation was built in at the project level; US\$20,000 had been allocated for this purpose out of a total budget of US\$500,000. In Viet Nam, Ministries received support in producing PM&E tools, and in Ghana the Country Office provided technical and financial inputs for a review of the health sector response to maternal mortality carried out by the Ministry of Health. A brief review of other interventions aimed at supporting national evaluation capacity (e.g. World Bank 2004) show that these are not systematic.²⁰ Staff at both Country Office and HQ levels noted that the central role of the UN should be to support the building of national capacity. Questions regarding support to national capacity were asked to 57 interviewees. Of these, 31 respondents thought that this should be a future priority, with no contrary views.

20. The UNDP Evaluation Office noted that 25 per cent of all Country Offices are involved in supporting PM&E capacity, and noted a number of other individual initiatives such as support to the Sri Lankan Evaluation Society. The UNICEF Evaluation Office noted that evaluation capacity building is a core theme for the agency; much of its work is based on developing competencies in this area.

4. Quality of evaluation resources

This Section reviews the quality and use of resources provided by HQ to support UNFPA evaluation exercises, in particular the Policies and Procedure Manual, the Programme Manager's Planning, Monitoring and Evaluation Toolkit, and the Regional Dialogues. Findings are based on a review of these documents and interviews.

Country Office respondents were unanimous in their praise for the HQ led Regional Dialogues on Results-based Planning, Monitoring and Evaluation. They found the following useful:

- ❑ The chance to interact with peers from other Country Offices in the region;
- ❑ Discussion of concepts of strategic planning and RBM; and
- ❑ Discussion of effective evaluations.

Country Office staff requested further dialogues as well as systematic follow-up to these efforts to support further learning.

Other HQ supported initiatives received more mixed reviews from UNFPA staff. The *Policies and Procedure Manual* was not being used to guide evaluations. The *Programme Manager's Planning, Monitoring and Evaluation Toolkit* was not being used on a regular basis, and respondents requested more operational content and an easier to use format. Some Country Offices had translated these guidance documents into local languages to share with participants. As one HQ respondent put it: "DOS needs to sell its products in a busy marketplace." Another said: "DOS products are useful, but not used." Of the 64 interviewees responding to questions on HQ resources, 31 agreed with the comments above, with three holding contrary views.

The message received from Country Office and HQ interviews was that UNFPA staff needs a range of options to support learning on PM&E. Some preferred on-line training and resources, others did not have good Internet access. There were two constants: Respondents valued face-to-face contact and peer education, and wanted easy to use tools and guidance.

5. Conclusions

From the total sample of 60 evaluations rated against the EQA matrix, 34 per cent scored satisfactory or better across all rating criteria. Against nine evaluation criteria specified in the 1997 *Policies and Procedures Manual* the sample fares better, with 49 per cent rated as satisfactory or better. Given that a 'satisfactory' rating is a minimal performance, our conclusion is that evaluation quality at UNFPA is currently unsatisfactory, and the evaluation function lacks credibility. Weak performance across regions suggests that weak evaluation performance is an organizational issue. Senior management (including the Executive Director, Deputy Executive Directors and Heads of Geographical Divisions) have reason to question whether they have adequate and trustworthy evaluative information on results, either to demonstrate UNFPA performance, or to improve programmes. However, good practices in almost all evaluative areas could be identified in a minority of cases.

Reasons for this unsatisfactory performance were revealed in interviews with 173 UNFPA staff, government counterparts and beneficiaries. UNFPA lacks a strong evaluation culture; that is, evaluations do not have a strategic role in decision-making. UNFPA does not have an adequate evaluation infrastructure or planning mechanisms. Across UNFPA, evaluations are currently being commissioned on an ad hoc basis, with little follow-up on recommendations and few possibilities to utilise lessons learned. The evaluation approach at UNFPA was perceived by staff and counterparts as being too accountability focused. Evaluations tend to be non-participatory, and beneficiary participation is rare, with 82 per cent of evaluations rated as 'missing' on the criteria "extent of stakeholders/beneficiaries involvement".

However, there was consistent support across UNFPA for a stronger evaluation function. No staff questioned the importance of evaluation or the central role it should play. Country Offices visited have been innovative and produced useful evaluations which have reshaped programming to make it more effective. In five out of seven good practice case studies savings or likely savings in programming were higher than the cost of the evaluation. These elements suggest there is a firm base upon which to build better evaluation practice.

UNFPA evaluations provide limited information on areas of current strategic importance to the organisation. An analysis of six criteria - assessment of capacity development, cultural issues, national ownership, partnership and alliance building, RBM, and the extent to which the evaluation enables UNFPA to engage in evidence based policy dialogue - revealed that attention to these criteria was satisfactory or better in 38 per cent of cases. Consequently, UNFPA is not currently positioned to answer the evaluative questions posed by the *Strategic Direction Positioning Statement*. These are major gaps that will need to be filled. Evaluation of gender mainstreaming and human rights were also found to be unsatisfactory.

Forty six per cent of evaluations in the sample were carried out solely by national evaluators, and a further 30 per cent by a mixed team of national and international evaluators, with the remaining 24 per cent carried out by international evaluators and/or CST advisers. Higher quality evaluations appeared to be produced by mixed teams of national and international evaluators. Key factors for improving the quality of UNFPA evaluations is for Country Offices to access high quality national evaluators; and to promote mixed teams consisting of national and international evaluators. Currently, an inadequate amount of time is being allowed for evaluations. Hence, there is insufficient time to examine: The key evaluative questions raised in the *Strategic Direction Positioning Statement*; for facilitating a participatory process with counterparts; or for systematic consultations with beneficiaries.

HQ-led Regional Dialogues on Results-based Planning, Monitoring and Evaluation were consistently praised and considered useful by staff in order to: Interact with peers from other Country Offices in the region; discuss concepts of strategic planning and Resource-based Management (RBM); and examine effective evaluations. Other HQ supported initiatives received more mixed reviews. Staff found that the *Policies and Procedures Manual* and the *Programme Manager's Planning, Monitoring and Evaluation Toolkit* need to be more user-friendly.

6. Lessons learned

During the country case studies, the following replicable factors were identified which led to good practice:

- ❑ 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.
- ❑ Evaluations need to pay special attention to regional and socio-economic disparities, especially in terms of trying to find out how UNFPA supported interventions are making a difference in the lives of poor women. Current assessment of effectiveness in evaluations is lacking in this area.
- ❑ Country Offices need to develop appropriate skills necessary for evaluating policy dialogue.
- ❑ Terms of reference are needed that clearly define the purpose of the evaluation, use and users, methodology, and objectives and process; well defined terms of reference lead to higher quality evaluations.
- ❑ There is no substitute for committed, well trained staff, with access to appropriate resources.
- ❑ Country Office staff should have more exposure to utilization and other participatory evaluation techniques. One of the Appreciative Inquiry findings was that a participatory process that involves stakeholders in reflection, analysis and future action can build strong ownership vis-à-vis evaluation findings. In the Ghana and Lebanon case studies full involvement of partners in clarifying the scope of work and developing the terms of reference helped build counterpart ownership. The Nicaragua case study and TSD evaluation of the gender based violence programme demonstrated that beneficiaries can play an important evaluative role. For a participatory process to succeed there must be an understanding of local culture and preferably local languages, and an adequate timeframe.
- ❑ Programme design that plans for evaluation facilitates the conduct of high quality evaluations. For example, an adequate monitoring system should be in place which provides key data for evaluators, and baseline and end-line surveys should be included along with ongoing monitoring and self-assessment systems. This lesson came across in several case studies, particularly Mozambique and Lebanon. Evaluations are especially effective if their results feed into new planning and decision-making processes. Therefore, the dissemination and use of evaluation findings should be planned for and budgeted from the outset.

An additional factor leading to higher evaluation quality was strong capacity on the part of national partners. Though this may not exist in some countries, in the Lebanon and Viet Nam case studies, for instance, there is strong counterpart capacity, and a willingness to promote evaluation. In these cases a utilization-focused and participatory approach promote counterpart ownership of evaluation recommendations, and subsequent follow-up, even where evaluation findings are critical. The support of PM&E capacity at the central government level requires influencing the way larger government systems work, and can only be adequately achieved in the medium to long-term through a joint effort with other donors.

7. Recommendations and Management Response

A strategy for improving evaluation quality

Recommendations	Management Response
<p>1. UNFPA should develop, implement, and monitor a strategy for improving evaluation quality.</p> <p>This could include:</p>	<p>1. UNFPA will develop an evaluation policy that includes a strategy to improve evaluation quality. The policy will reflect current UNEG guidance and enhance conceptual clarity on the role and functionality of evaluation vis-à-vis other analytical functions and assessments in UNFPA. The policy will also provide broad principles which will govern future evaluation work and provide clarity with regard to management accountabilities of various organizational units for undertaking and using evaluations.</p>
<p><input type="checkbox"/> Updating the <i>Policies and Procedures Manual</i> section on monitoring and evaluation, following UNEG guidance.²¹</p>	<p><input type="checkbox"/> UNFPA will undertake all the necessary revisions to the PPM to ensure proper compliance with UNEG standards.</p>
<p><input type="checkbox"/> Determining what levels of improvement in evaluation quality is required, and over what time period. Meeting these targets could be measured by a full-scale meta-evaluation, similar to the current meta-evaluation, in 2010. There could also be periodic smaller-scale assessments of quality - e.g. at six month periods.</p>	<p><input type="checkbox"/> UNFPA will establish targets for improving evaluation quality over specific time frames. A quality control and monitoring function will also be established to measure progress at periodic intervals. The UNFPA Balanced Scorecard can also be used to monitor progress at the Executive Committee level. A second meta evaluation using the same assessment matrix as the 2005 EQA will be undertaken in 2010.</p>
<p><input type="checkbox"/> Providing Country Offices and CSTs with training in participatory and utilization focused evaluation approaches. This could highlight different entry points for participation (e.g. development of ToRs, workshops on results); and different methods that can be used to promote participation of stakeholders, including counterparts and beneficiaries</p>	<p><input type="checkbox"/> UNFPA will continue investing in relevant training and knowledge sharing and will further develop the approach.</p>
<p><input type="checkbox"/> Hiring additional PM&E staff. These staff could be located at the regional (e.g. CST) or Country Office level. However, UNFPA should examine the possibility of establishing focal points for particular sectors before deciding on this issue. If new staff are hired as PM&E focal points, this should not be a substitute for mainstreaming PM&E throughout the organisation.</p>	<p><input type="checkbox"/> Under the UNFPA regionalization initiative, four additional M&E advisor posts to be located within the regional offices will be included as a means of creating additional capacity in the area of M&E and to assist with mainstreaming efforts. In this connection, UNFPA will develop a competency framework for M&E Officers at all levels in the Fund and corresponding job descriptions in line with UNEG practice.</p>
<p><input type="checkbox"/> Integrating monitoring and evaluation in core competencies of relevant staff. For example, evaluation could be included in the Analytic and Strategic Thinking, and Results Orientation competencies.</p>	<p><input type="checkbox"/> UNFPA will amend the relevant competencies as recommended.</p>
<p><input type="checkbox"/> The EQA matrix could be adapted to a set of quality standards that could be attached to the evaluation terms of reference.</p>	<p><input type="checkbox"/> UNFPA will develop a set of minimum evaluation quality standards which will be attached to all evaluation Terms of Reference.</p>
<p><input type="checkbox"/> Providing guidance on the optimal length of time for different kinds of evaluations.</p>	<p><input type="checkbox"/> UNFPA will include these details within the toolkit.</p>

A strategy for improving evaluation quality (continued)

<input type="checkbox"/> Continuing to invest in Regional Dialogues.	<input type="checkbox"/> UNFPA will continue to invest in training of staff in results-based management.
<input type="checkbox"/> Maintaining a searchable electronic database of evaluations. This would require quality screening of evaluations against key criteria in the EQA assessment matrix to determine acceptance in the database (e.g. completeness of report, quality of methodology, attention to key cross-cutting themes, and OECD-DAC criteria). The database should be staffed with a full-time manager, as databases not adequately managed are not used. The data base manager could also support future meta-evaluation activity.	<input type="checkbox"/> UNFPA will create a searchable database of quality-controlled evaluations which will be available to the public. The database will be managed by DOS who may make use of the services of consultants to manage the database.

Follow-up to recommendations

Recommendation	Management Response
<p>2. All evaluations should include an action plan for recommendation follow-up.</p> <p>This should include:</p>	<p>2. UNFPA's new evaluation policy will make managers accountable for using evaluations and for building an evaluation culture. The new minimum standards will also require planned evaluation uses and users to be identified. UNFPA will verify adherence to these procedures during its routine oversight work.</p>
<p><input type="checkbox"/> Response to recommendations by relevant parties, prioritization of recommendations, and a recommendations tracking matrix. COARs should report on management attention to follow-up to evaluation recommendations.</p>	<p><input type="checkbox"/> As part of the new evaluation policy, all managers undertaking or commissioning evaluations will be responsible for preparing an action plan for follow up and will be held accountable for doing so. The use of a recommendations tracking matrix will be encouraged as a best practice. A mechanism for reporting on evaluation follow-up status will be devised, and this will be spot-checked during the oversight missions. Responsibility for conducting and following up on evaluations will also be reflected in the balanced scorecards.</p>
<p><input type="checkbox"/> Evaluations should be expected to identify clearly use and users in terms of reference and reports.</p>	<p><input type="checkbox"/> This aspect will be included as part of the new evaluation minimum standards.</p>
<p><input type="checkbox"/> The Executive Summary should be translated into local languages.</p>	<p><input type="checkbox"/> This suggestion will be included in the toolkit.</p>

Evaluating current strategic priorities

Recommendation	Management Response
<p>3. Guidance and staff capacity development is needed for evaluation of current strategic priorities.</p> <p>This should include guidance on evaluating participation in Sector Wide Approaches.</p>	<p>3. UNFPA is currently developing the new strategic plan (SP) for the period 2008-2011. This includes programme focus areas, capacity building strategies and crosscutting programme principles. The SP will provide the basis for developing guidance on methodologies for evaluating UNFPA performance in achieving results and applying strategies and principles. The guidance will build upon the national capacity development strategy being developed by a Task Team as well as other substantive policies and strategies also developed by TSD.</p>

Gender mainstreaming and human rights

Recommendations	Management Response
<p>4. UNFPA needs to demonstrate a measurable improvement in attention to human rights and gender equality in its evaluations. It can do this by, <i>inter alia</i>:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Ensuring that terms of reference include promoting human rights and gender equality as areas to assess. <input type="checkbox"/> Ensuring that methodologies include details on how contributions to human rights and gender equality will be measured. <input type="checkbox"/> Ensuring that data in evaluation reports is sex-disaggregated. <input type="checkbox"/> Engaging evaluators with a good understanding of the importance of human rights and gender mainstreaming. <input type="checkbox"/> Ensuring that there are separate sections on human rights and gender equality in the evaluation, and that they are also covered in all other main sections of the evaluation as a cross-cutting theme. <input type="checkbox"/> Ensuring that at least one recommendation deals directly with human rights, and one with gender equality. 	<p>4. The new minimum evaluation standards will place an emphasis on improving evaluation coverage of issues related to gender equality and human rights.</p> <p>In order to be able to evaluate, it is important that from the very beginning UNFPA interventions clearly establish how gender equality and human rights issues will be addressed and measured including the selection of appropriate indicators identified in the results framework. Gender mainstreaming needs to be clearly defined and the results of the Focusing on Gender report will also be incorporated.</p> <p>Depending on the focus of the programme and the scope of the evaluation, human rights and gender equality will be covered and related recommendations included, wherever possible.</p>

National capacity

Recommendations	Management Response
<p>5. A long-term strategy for UNFPA is to support, with other UN agencies, national evaluation capacity.</p>	<p>5. UNFPA will include in its evaluation policy recognition of the fact that developing national evaluation capacity and building effective partnerships for evaluation represent important strategic objectives of UNFPA in its broader effort to promote ICPD.</p> <p>M&E officers will also have this aspect included in their ToRs</p>
<p>6. Mixed teams of national and international evaluators should be used where feasible.</p>	<p>6. UNFPA's M&E toolkit will be revised to reflect the desirability of having mixed teams of international and national evaluators on evaluation teams.</p>

Methodologies

Recommendation	Management Response
<p>7. Planned method sections of evaluations could be screened (e.g. against an adapted set of EQA standards - see Recommendation 1). Evaluations over a certain budget limit, and/or a random sample of evaluations, could have a methodological review outside of the Country Office.</p>	<p>7. In the case of large-scale evaluations with planned budgets in excess of US\$100,000, UNFPA will screen all evaluation methodologies for quality.</p>

Completeness of reports

Recommendation	Management Response
<p>8. Screening of terms of reference could be carried out by Geographical Divisions, with spot-checking by DOS. A quality checklist for terms of reference could be developed, building on the guidance in the <i>Programme Manager's Planning, Monitoring and Evaluation Toolkit</i>.</p>	<p>8. UNFPA will revise the PPM and the M&E toolkit on evaluation accordingly to reflect the need for improved oversight and quality control of ToR preparation. Checking on the quality of ToRs will also become a feature of UNFPA's regular oversight work. The regionalization initiative will also include regional M&E Advisors which will enhance capacity for spot-checking ToRs</p>
<p>9. A checklist based on the model Table of Contents in the <i>Programme Manager's Planning, Monitoring and Evaluation Toolkit</i> could be developed for screening of Tables of Contents.</p>	<p>9. DOS will develop such a checklist and include as part of the toolkit.</p>

HQ guidance

Recommendations	Management Response
<p>10. HQ resources should be edited to make them more user-friendly. A professional editor should be hired for this purpose.</p>	<p>10. DOS will review the M&E toolkit and make adjustments as required.</p>

Bibliography

ALNAP (2004) *Review of Humanitarian Action*. London: Action Learning Network for Accountability and Performance in Humanitarian Action.

Bamberger, M. (ed)(2000) *Integrating Quantitative and Qualitative Research in Development Projects*. Washington D.C., World Bank.

CARE (2001) *The MEGA Evaluation: A Review of Findings and Methodological Lessons from CARE Final Evaluations*. Atlanta: CARE USA Program Division.

CARE (2003) *The MEGA 2002 Evaluation: A Review of Findings and Methodological Lessons CARE Final Evaluations, 2001-2002*. Atlanta: CARE USA Program Division.

Cooperrider, D. and D. Whitney. (2002) *Appreciative Inquiry: The Handbook*. Euclid: Lakeshore Publishers.

Cooperrider, D.L. and D. Whitney. (1999a) 'Appreciative Inquiry' in *Collaborating for Change*. Ed. P. Holman and T. Devane. San Francisco: Berrett-Koehler Publishers.

Cooperrider D.L. and Whitney, D. (1999b) 'Appreciative Inquiry: A Positive Revolution in Change' in *The Change Handbook*. Ed. by P. Holman and T. Devane. Berrett-Koehler Publishers.

Consulting and Audit Canada (2004) *Strategy and Implementation Plan for Results-Based Management and Results-Based Budgeting: October 2004*. Consulting and Audit Canada.

DANIDA (2004) *Metaevaluation: Private and Business Sector Development*. Copenhagen: Danish Ministry of Foreign Affairs.

Ginsburg and Rhett (2003) 'Building a Better body of Evidence: New Opportunities to Strengthen Evaluation Utilization.' *American Journal of Evaluation* 24 (4), 489-98.

Hammond, S. (1998) *The Thin Book Of Appreciative Inquiry*. Plano, Texas: Thin Book.

Hansen, K. (2004) 'Appreciative Inquiry and Evaluation Part 2'. *Criterion, spring*.

Henry, G. (2003) 'Influential Evaluations.' *American Journal of Evaluation* 24 (4), 512-24.

IDRC (2004) *Annual Report on Evaluation Findings*. Ottawa: International Development Research Centre.

Kanbur, R. (ed)(2003) *Q-Squared : Combining Qualitative and Quantitative Methods in Poverty Appraisal*. Delhi: Permanent Black.

Kirkpatrick, D. (1998) *Evaluating Training Programs: The Four Levels*. San Francisco: Berrett-Koehler Publishers.

- Lipsey, M. (2000) 'Meta-Analysis and the Learning Curve in Evaluation Practice' in *American Journal of Evaluation*, 21 (2), pp207-213.
- Ministry of Foreign Affairs, The Netherlands (2004) *Quality Assessment of Evaluations completed in 2002, Final Report: August 2004*. The Hague: Policy and Operations Evaluation Department.
- Patel, M. (2002) 'A meta-evaluation, or quality assessment, of the evaluations in this issue, based on the African Evaluation Guidelines: 2002.' *Evaluation and Program Planning* 25, pp329-332.
- Patton, M. (2003) 'Inquiry into Appreciative Evaluation.' *New Directions in Evaluation* 100, 85-98.
- Patton, M. (2002) *Qualitative Research and Evaluation Methods*. Thousand Oaks: Sage Publications.
- Patton, M. (1997) *Utilization-focused Evaluation: The New Century Text*. Thousand Oaks: Sage Publications.
- Preskill, H. (2001) *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance & Change*. London: Sage.
- Schwartz, R. & Mayne, J. (2005) 'Assuring the quality of evaluative information: theory and practice' in *Evaluation and Program Planning*, 28 (1), pp1-14.
- Scott-Little, C., Hamann, M.S. & Jurs, S. (2002) 'Evaluations of After-School Programs: A Meta-Evaluation of Methodologies and Narrative Synthesis of Findings' in *American Journal of Evaluation*, 23 (4), pp387-419.
- Stufflebeam, D. (2001) 'Guiding Principles Checklist for Evaluating Evaluations'. Mimeo.
- Stufflebeam, D. (2001a) 'The Metaevaluation Imperative.' *American Journal of Evaluation* 22 (2), pp183-210.
- Swiss Agency for Development and Cooperation (2004) *Quality Assessment of SDC's External Evaluation Reports: Berne, November 2004*. Swiss Agency for Development and Cooperation.
- Tingle, L, DeSimone M. & Covington, B. (2003) 'A Meta-Evaluation of 11 School-Based Smoking Prevention Programs' in *Journal School of Health*, 73 (2), pp.64-67.
- UN (1999) *Key actions for the further implementation of the programme of Action of the International Conference on Population and Development*. New York: A/RES/S-21/2.
- UN (1995) *Report on the International Conference on Population and Development*. New York: A/CONF.171/13/Rev.1.
- UNEG (2005) *Norms for Evaluation in the UN System*. New York: United Nations Evaluation Group.
- UNFPA (2005) *Report of the Executive Director for 2004: Progress in Implementing the Multi-Year Funding Framework, 2004-2007*. New York: UNFPA DP/FPA/2005/7 (Part 1)

UNFPA (2004) *Programme Manager's Planning Monitoring and Evaluation Toolkit*. UNFPA, Division for Oversight Services.

UNFPA (2004a) *Policies and Procedures: Country Programme Monitoring and Evaluation*. UNFPA, mimeo.

UNFPA (2004b) *DOS/SPO/CST/CO Dialogue on Results based Planning, Monitoring and Evaluation: Addis Ababa, 1-4 November 2004*. UNFPA, mimeo.

UNFPA (2004c) *DOS/SPO/CST/CO Dialogue on Results based Planning, Monitoring and Evaluation: Bratislava, 6-9 July 2004*. UNFPA, mimeo.

UNFPA (2004d) *DOS/SPO/CST/CO Dialogue on Results based Planning, Monitoring and Evaluation: Harare, April 2004*. UNFPA, mimeo.

UNFPA (2004e) *DOS/SPO/CST/CO Dialogue on Results based Planning, Monitoring and Evaluation: Kathmandu, June 2004*. UNFPA, mimeo.

UNFPA (2004f) *DOS/SPO/CST/CO Dialogue on Results based Planning, Monitoring and Evaluation: Bratislava, 6-9 July 2004*. UNFPA, mimeo.

UNFPA (2004g) *The Multi-Year Funding Framework 2004-2007. Report of the Executive Director*. New York, DR/FPA/2004/4.

UNFPA (2004h) *Periodic Report on Evaluation*. New York: DP/FPA/2004/12.

UNFPA (2004i) *Results-Based Programme Planning Monitoring and Evaluation Checklist*. UNFPA: Division for Oversight Services.

UNFPA (2003) *GD/SPO/DOS Dialogue on Results-based Planning, Monitoring and Evaluation: Status Report as of August 2003*. UNFPA. Mimeo.

UNFPA (2001) *Report for the Executive Director for 2000: The Multi-Year Funding Framework*. New York: DP/FPA/2001/4/Part II.

UNFPA. nd. *CPAP Quality Assurance Checklist*. UNFPA, mimeo.

UNFPA. nd. *UNFPA Policy Statement on Results Based Management*. UNFPA, mimeo.

UNICEF (2004) *UNICEF Evaluation Report Standards*. New York: UNICEF.

UNICEF (2003) *The Quality of Evaluations Supported by UNICEF Country Offices, 2000-2001*. New York: UNICEF.

USAID (2001) *Evaluation of Recent USAID Evaluation Experience*. Washington DC, Center for Development Information and Evaluation, Working Paper 320.

World Bank (2005) *Influential Evaluations: Detailed Case Studies: January 2005*. Washington, D.C.: The World Bank Operations Evaluation Department.

World Bank (2004) *Evaluation Capacity Development: OED Self-Evaluation: June 2004*. Washington, D.C.: The World Bank Operations Evaluation Department.

Annex 1

Terms of Reference

BACKGROUND

Within UNFPA, programme evaluations are undertaken in a decentralized manner. In recent questionnaire surveys conducted as part of the Country Office Annual Reporting (COAR) exercises, UNFPA Country Offices and CSTs reported some 181 evaluations or evaluative reports undertaken in 2003 and 151 in 2002.¹ The Fund therefore produces a substantial body of evaluation work each year, which is intended to form the bedrock of the Fund's efforts to develop and improve its programmes and to demonstrate results.

The Oversight and Evaluation Branch (OEB) has been concerned with building a stronger evaluation culture within the Fund and developing its quality-assurance, control and monitoring functions in relation to evaluation. This role is anticipated in the ToR for the Division of Oversight Services (DOS), which, among other things, suggests that DOS "monitors and advises on organizational performance, policies, procedures and internal control systems to promote economy, effectiveness, efficiency and integrity in UNFPA operations. It sets organizational standards for monitoring and evaluation and promotes a results-based monitoring and evaluation culture within the organization". Specific tasks for the Division include providing "advice on monitoring and evaluation processes" and promoting "the use of lessons learned from evaluations to facilitate corporate learning". OEB is also specifically tasked with making recommendations for improving organizational performance based on the findings of other oversight and evaluation activities.² Notable efforts to improve the quality of evaluation include production of the 1997 M&E guidance materials, the development of an M&E toolkit including several evaluation-related modules, and most recently, the new results-based M&E guidelines. The 2003 review of regional programmes similarly made recommendations regarding the need to improve evaluation quality.

As part of its regular oversight function, the Division now considers it strategic and useful to undertake a meta-evaluation type quality assessment of UNFPA's recent evaluation work both as a means of identifying lessons learned and bringing about further quality improvements. This proposal builds on current trends in the evaluation field, which increasingly recognize meta-evaluation as an effective tool for organizational learning about the role of evaluation in improved programming.³

¹ These figures do not include inter-regional and regional evaluations, a number of which have also been undertaken.

² See DOS/OEB ToR dated 21 January 2003.

³ Mark W. Lipsey, Meta-Analysis and the Learning Curve in Evaluation Practice. *The American Journal of Evaluation*, Volume 21, Number 2, Spring Summer 2000, p. 211.

OBJECTIVES

The EQA has four major objectives:

Short-term:

1. To support organizational learning about effective evaluation practice.
2. To highlight effective evaluation practices that can be replicated
3. To establish a baseline for future quality assessments

Longer-term:

4. To produce more consistent and better quality evaluations

KEY ISSUES

The proposal to undertake a systematic quality assessment of the Fund's evaluation work represents both a timely and relevant endeavour. To help the Fund contribute more effectively to the implementation of the ICPD Programme of Action and to the Millennium Development Goals, a set of new Strategic Directions (SD) has been put in place. The new Strategic Directions place considerable emphasis on building UNFPA's credibility on ICPD follow up issues through the use of evidence-based arguments, the dissemination of effective programmatic approaches, including lessons learned and replicable best practices. In this context, evaluation quality must be seen as a matter of utmost concern.

The concern for quality encompasses a broad range of issues. These include not only matters of methodological appropriateness, analytical rigor and adequate report content but also extend into utility issues such as management's capacity to commission and manage good evaluations and then to utilize the findings, conclusions and recommendations. Transferring learning from specific evaluation settings to a wider range of programme settings represents an additional concern.

Pursuant to the above, the following key concerns guide the EQA:

Quality of evaluation reports

- In light of current organizational goals and priorities, are UNFPA evaluations presently evaluating the "right" kinds of things in the "right" kinds of ways? How well presented are the evaluation results?
- How do current approaches found within a broad sample of UNFPA evaluation reports compare with internationally-recognized best practices in evaluation? How well do these reports compare against the basic evaluation requirements set down in the 1997 policies and procedures guidelines?
- Are UNFPA's current evaluation practices and approaches adequate and sufficient to help fulfill the requirements of UNFPA's new Strategic Directions? In light of the MYFF, are there currently any gaps in evaluation coverage in relation to key strategic themes or issues?
- What improvements are needed to support more consistent and better quality evaluations and to strengthen their role in implementing the Strategic Directions?

Quality of evaluation management and attention to follow up

- Is UNFPA able to commission useful and timely evaluations, which address relevant issues and concerns for programme managers? Do staff and partners have a clear understanding of evaluation requirements?
- Are UNFPA evaluation results including lessons learned effectively being used for their primary intended purposes, namely of supporting programme modification and improvement and of providing accountability for results? How useful do managers find evaluation results?
- Do UNFPA evaluations support and influence policy and policy dialogues? Do evaluations support the development of evidence-based arguments and help support the broader UNFPA policy agenda at the national and international level? Are they designed to do so?
- Does UNFPA management provide systematic attention to evaluation results? How widely and to whom are evaluations circulated and for what ends? Are management accountabilities clear?
- In relation to all points above, what gaps currently exist and what improvements might potentially be considered to make evaluations more useful, especially as regards their role in supporting implementation of the Strategic Directions?

Quality of available resources

- Are available human and financial resources sufficient for supporting evaluation work? What gaps exist?
- What barriers and opportunities do staff and other relevant stakeholders perceive in the effort to produce better and more consistent evaluations?
- How clear and relevant are the available guidance materials on evaluation? Do they support implementation of the new Strategic Directions? Are evaluation stakeholders and UNFPA staff able to make effective use of the guidance materials?

Identification of quality evaluation practices

- What instructive examples of quality evaluation results and practices can be highlighted to guide future evaluation efforts? In this context it is important to provide analytical examples, which illustrate or exemplify the application of key evaluation criteria and standards within the UNFPA context and provide successful examples of effective evaluation processes and management use, especially in cases where evaluations were used to support policy development and dialogue. What factors seem to be associated with these success stories?

Synthesis of evaluation findings (Optional key issue depending on findings and availability of a critical mass of good quality evaluations)

- What key findings and lessons can be distilled from the reports studied in relation to organizational programme priorities (i.e. Human rights-based approach, gender mainstreaming, people-centered planning, national capacity development, culturally sensitive programming, partnership and alliance building and implementation of results-based management)?

INTENDED USES & USERS

The basic findings, recommendations and conclusions of the EQA are intended in the short term to be used to support organizational learning about what effective evaluation practice looks like within UNFPA. The recommendations may also be used to develop a comprehensive longer-term follow up plan aimed at further consolidating evaluation quality within the Fund.

Although the assessment of written reports will be carried out largely in relation to current internationally recognized best practices, the assessment will need to interpret findings and results in this same light. Data in relation to how well the sampled evaluations as a whole meet the minimum requirements set out the 1997 evaluation guidelines will be included.

A number of other specific uses have also been identified:

- The EQA results can assist UNFPA Staff and national counterparts to identify useful and concrete examples of effective evaluation practice, which may in turn serve as examples to guide their own evaluation efforts. Successful practices may also be integrated into future PM&E training and other evaluation capacity building events.
- The EQA will help establish priority actions for improvement, possibly through both the development of improved guidelines and/or policy frameworks. Specific actionable items in this regard could potentially include expansions and/or enhancements to the M&E toolkit, development of advocacy points for improved strategic directions in relation to PM&E and the identification and dissemination of lessons from pilot interventions, which can help support policy dialogues.
- The EQA may also potentially evolve into an on-going quality assessment process. Seen against a longer-term backdrop, this initial EQA will serve as a baseline study, or stock taking exercise against which potential future quality assessments could be compared and monitored.

METHODOLOGY

The UNFPA glossary of planning, monitoring and evaluation terminology defines a meta-evaluation as “a type of evaluation that aggregates findings from a series of evaluations. Also, an evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators”.

In order to strike an appropriate balance between methodological rigor and organizational learning, a blend of both meta and participatory evaluation techniques will be adopted. Building in participation will not only help reinforce organizational learning, it will also offer opportunities for methodological triangulation. Appreciative Inquiry (AI) will also be adopted as a further methodological compliment. AI has the potential to increase the usefulness of findings and help build momentum towards greater participation and follow up action in the longer-term.

Mixed method approaches are by now more or less now considered “best” practice. Their adoption for this specific exercise builds on lessons learned from recent meta- evaluations undertaken by UNICEF, IDRC and ALNAP. All these exercises have found that an exclusive focus on assessing evaluation quality through written reports misses important aspects of evaluation quality, especially those related to evaluation’s learning effects – also known as process use – and stakeholder utility issues. In this context, the 2003 IDRC meta evaluation expresses concern that “looking exclusively at evaluation reports...can under-represent the true quality of evaluation. Evaluation reports do not always provide a full description of evaluation processes and procedures, and as a result the (meta-evaluation) system will sometimes fail to register positive scores on indicators of quality when evaluators may have employed sound evaluation processes”.⁴

⁴ IDRC. Annual Report on Evaluation Findings 2003. Ottawa, Canada, p.23

UNICEF's own meta evaluators similarly lamented the heavy reliance on assessing the quality of evaluation work primarily through written reports. They concluded that such an approach constitutes "a narrow base for our recommendations about how evaluation quality might be improved".¹

ALNAP, the evaluation and learning network within the humanitarian aid world, has now begun building in more interaction between meta-evaluators and evaluation managers during its annual meta evaluation exercise. They note in this regard "This shift has in turn brought about a greater focus on evaluation process issues and the link with quality"²

UNFPA's first EQA exercise can benefit from these experiences by adopting a balanced approach. On the one hand, the exercise needs to have solid methodological foundations for building actionable, appropriate and credible recommendations in the short to medium term; on the other hand, the exercise should also attempt to lay the foundations for a longer-term evaluation quality improvement process throughout the entire organization.

SCOPE OF ASSESSMENT

The EQA will focus on a sample of evaluation reports managed by UNFPA Country Offices, regional evaluations as well as evaluations managed by TSD.

The EQA will be structured through a number of phases as follows:

1. EQA Design
2. Desk Assessment and Case Studies
3. Report Preparation
4. Validation and Dissemination

Design of Methodology

A methodological approach paper (MAP) will initially be prepared and submitted for review by a Reference Group. The paper will, among other things, specify the general indicators to be used during the Desk Assessment phase and details regarding the design of the case studies.

Desk Assessment

The Desk Assessment will cover a sample of:

1. Country office-managed evaluations (including project, thematic, sub programme as well as country programme evaluations)
2. Regional evaluations (including both project and thematic); and
3. Inter-regional and inter-country evaluations.

A sample of up to 80 evaluation reports will be identified making use of the 2002 and 2003 evaluation reports contained in the Country Office Annual Reports (COARs). Regional and inter-regional evaluations will be obtained separately. As part of the sampling procedure, Country Offices and CSTs will be requested to forward to the EQA samples of evaluations, which they found to be useful. Each evaluation report will be assessed following a common set of indicators and with reference to a standard rating guide. One external reader and one DOS Shadow Reader will review and rate each report.

¹ UNICEF. The Quality of Evaluation Reports Supported by UNICEF, 2000-2001.

² ALNAP Review of Humanitarian Action in 2003. Field Level Learning. Overseas Development Institute, London, UK.

Case Studies

Subsequent to the Desk Assessment, up to seven case studies involving a number of field visits will be undertaken to help add balance and texture to the report and to follow up on key issues for which answers may not readily be obtained from the desk assessment alone. These include aspects of evaluation quality related to management use. Each case study will be conducted by two team members and other DOS staff members as and when needed.

Use of Questionnaire

A brief questionnaire and telephone interviews may also be included in order to further compliment the findings of the seven individual case studies and to explore key issues among a broader group of stakeholders including EVALNET members (if possible).

PRODUCTS

The EQA will result in the following outputs:

A methodological approach paper (MAP): the paper will outline the methodology to be adopted for the meta evaluation including the quality assessment grid for guiding the desk review, an evaluation matrix, details on the methodology for the selection of the good practice case studies. The MAP also includes the TORs for each case study to ensure that all case studies follow a similar approach, ask similar questions and follow a common format and table of contents for presentation. **Deadline: 20 January 2005**

An aide-mémoire: once both the desk review and the field work have been completed, a brief aide-mémoire will be prepared outlining key evaluation findings, tentative conclusions and potential recommendations. **Deadline: June 17 2005**

A first draft of a full EQA report: the evaluation report consists of a main report synthesizing analytical highlights, key findings, conclusions and recommendations in relation to all the main issues. The table of contents and approximate chapter lengths will be finalized in a logical manner consistent with ensuring complete coverage of all the key analytical issues listed in these ToR. The report will also include a number of annexes including i) the final set of ToRs; ii) summary tables showing the aggregated results for each quality assessment standard used for the Desk Review; iii) separate annexes summarizing the results of each case study; iv) a list of reference materials used. **Deadline: 28 July 2005**

A final EQA report: the final report will reflect all of the written comments received and all other feedback obtained from any meetings or workshops arising from the circulation of the first draft report. **Deadline: 1 October 2005**

All of the above written outputs will be prepared in **English** and will reflect the collective efforts of the team members. Adequate capacity must exist within the team as a whole to review reports and to conduct case studies in three working languages: English, French and Spanish. Although not every team member will be required to speak and/or to read all three, the evaluation team will be composed in a way which ensures that all the languages can be covered within the team.

The full evaluation report will mainstream and highlight as appropriate examples of good UNFPA evaluation practice in a manner that attracts the reader's attention. All judgments will be supported by summary data. Quantitative data will be presented in a user-friendly way including easy-to-read and understand tables and graphs. Qualitative data will be presented in a similar vein and include comments from a wide range of actual interviewees using their own words and in a manner which is

illuminating for the reader. The report will also make extensive use of shaded inset boxes highlighting examples of effective evaluation practice, lessons and/or instructive experiences.

All recommendations must be logically derived from and consistent with the report's main findings and conclusions. Recommendations will be formulated in such a way as to avoid convoluted language, and in this way formulated in as specific, practical and operational a manner as possible from the point of view of the intended user. Recommendations will also be clearly prioritized according to different categories of organizational priority, namely i) top urgent; ii) important and/or iii) desirable. In this context, a final EQA workshop will assist with building appropriate and actionable recommendations.

WORK PHASES & PROCESSES

A Work Plan detailing the different EQA work phases and processes is annexed to these ToR. This plan may need to be adjusted and modified as work progresses subject to the agreement of team members and the Evaluation Manager.

TEAM COMPOSITION & REQUIREMENTS

The EQA team will consist of three members:

A Team Leader: he/she will be an external consultant and will be responsible for taking the lead in producing all the written outputs specified above in consultation with and based upon written inputs from the other team members. He/she will also be responsible for assigning clearly defined roles and tasks to each team member in a manner as to ensure that all the required evaluation outputs are produced on time. The Team Leader will also be expected to research and to write up four separate case studies.

The Team Leader must possess a good grasp of evaluation methodology and extensive hands-on experience with conducting applied evaluations in an international setting similar to that in which UNFPA operates.

Other qualities for the Team Leader include prior experience with conducting meta evaluations of international development aid co-operation and some knowledge of relevant UNFPA programming areas and issues. Knowledge and practical experience with planning and conducting participatory evaluations would also be highly desirable.

An estimated **87** non-consecutive workdays will be required on the part of the Team Leader to produce the required outputs as per the work plan. This does not include any optional work in relation to undertaking an evaluation synthesis as identified under the key issues section.

A Second International Team Member: The second evaluator will also be an equally senior international evaluation consultant who is highly experienced in participatory evaluation processes and techniques as well as the production of evaluation guidance materials within large international organizations.

He/she will be required to provide written and verbal inputs to the Team Leader for the development of the methodological approach paper (MAP), especially as this relates to the mainstreaming of appreciative inquiry concepts and processes into the EQA. He/she will also be required to undertake assessments of evaluation reports following an established approach. He/she will also be required to research and write up following a standard methodology a minimum of three and possibly four

individual country case studies. The second team member will also be required to design and make use of participatory workshops and techniques throughout the EQA process.

An estimated **66** workdays will be required by the Second Team Member to provide the required inputs to the Team Leader.

A UNFPA Evaluation Advisor located in the Oversight and Evaluation Branch will be a full team member and will be responsible for the overall management of the EQA. This includes key tasks such as advising team members on UNFPA policies and procedures, bringing an internal organizational perspective to the team's internal discussions, facilitating the work of team members and their interactions with UNFPA staff and managers, co-ordinating logistical and travel arrangements and organizing key participatory events including Reference Groups (RG) meetings and workshops at New York Headquarters. He will participate in assessing evaluation reports as a "Shadow reader" and will also undertake field visits as required to support the case studies. He will also ensure the timely circulation of key draft documents to stakeholders as and when required. He will also participate in drafting parts of the final report as needed and for making a final presentation to the Programme Committee on key EQA findings and recommendations. The Advisor will devote an estimated **100** non-consecutive workdays to these tasks for both the analytical work involved in assessing the reports, going on field trips and managing the process at headquarters.

ROLES & RESPONSIBILITIES

Although the EQA team is expected to operate independently, an **EQA Reference Group** (RG) composed of select staff members from HQ and UNFPA field offices will be established by the Division of Oversight Services Oversight and Evaluation Branch (DOS/OEB) early on in the process.

The RG will review key outputs and to provide feedback for guiding the work of the assessment team. The RG will also serve to help build greater consensus around assessment methodologies, key findings, conclusions and any relevant follow up actions, which may be indicated.

RG feedback will be obtained through both face-to-face meetings at HQ as well as through e-mail contact and as appropriate telephone interviews with field-based members. The RG will meet in New York City in particular to i) provide comments and reactions to the proposed assessment methodology, contained in the methodological approach paper (MAP) ii) to review the aide-mémoire document and the initial evaluation results upon completion of the research phase and iii) to participate in a short workshop to help validate the EQA results and to contribute to conclusion drawing and recommendation formulation. At times, the RG may be expanded as appropriate to include other members.

The Geographical Divisions and TSD within NYHQ will be responsible for co-ordinating the selection of evaluation reports to be included in the Desk Assessment. To assist with the selection, they will make use of the existing lists prepared by DOS based on data extracted from the COARs from 2002 and 2003. To increase the sample size as much as possible, DOS may also include samples of recent evaluation reports which have come to its attention and of which it has obtained copies. The Geographic Divisions and TSD will also assist with the identification of case studies within their regions and for assisting with arrangements for team field visits as appropriate and within a time frame consistent with the attached Work Plan.

Individual UNFPA Country Offices **selected for case studies will be responsible to receive team members and to make available adequate staff time as well as any relevant internal records, reports and files, which the Team can make use of in developing its findings. They will also set up in advance meetings with government and key partners to discuss evaluation related issues.**

KEY DOCUMENTS

The main source documents will be the individual UNFPA evaluation reports. UNFPA's Geographical and Technical Support Divisions will be requested to co-ordinate the selection of a sample of evaluations for the EQA exercise.

In developing its analytical approach, the Evaluation team will be required to draw from and make reference to the existing literature on meta evaluation contained within evaluation journals. Special reference will also be made to the Joint Committee on Evaluation Standards. To the extent possible, references to lessons learned from other international agencies similar to UNFPA in using meta evaluations should be referred to and built upon. (e.g. UNICEF, IDRC, CARE, UNCDF, ALNAP, USAID etc...)

The evaluation team should also consult the following specific M&E related guidance and policy materials:

- The 1994 International Conference Population & Development Plan of Action
- Results-Based Planning, Monitoring and Evaluation Checklist
- The CPAP Checklist
- The CPAP and AWAP 2004 Guidelines
- The 1997 Programme Guidelines
- The UNFPA Programme Manager's Planning, Monitoring and Evaluation Toolkit
- The 2004 Monitoring & evaluation Guidelines
- The UNDAF Guidelines
- Dialogue Reports on results-based planning, monitoring and evaluation (PM&E)

Work Plan

PHASE	Key Products & Processes	PHASE	Key Products & Processes
1. Design		2. Research	
<p>January/February 2005</p> <p>Methodological Approach Paper</p> <p>The EQA team will develop a Methodological Approach Paper which will provide full details on the approach to be used. This should include the sampling strategy and the proposed assessment grid. A simple evaluation matrix relating data collection methods to core questions should also be included. The approach paper will further detail the interface between the meta desk assessment and the participatory processes. The participatory component will also explore the possibility of using Appreciative Inquiry (AI) principles during these interactions as a base for making results more digestible to all stakeholders. AI essentially involves identifying the best of what an organization is currently doing and then seeks out ideas and suggestions on how the organization can replicate these successes elsewhere.</p> <p>Draft Submission Deadline: 20 January 2005</p> <p>Reference Group Review Meeting</p> <p>The Draft Approach Paper will be presented to the RG prior to its finalization. The Approach paper is then finalized.</p> <p>Total Estimated Time: 9 working days by Team Leader and 9 working days by Team Members. Evaluation Advisor inputs as needed from New York</p> <p>Finalization of methodology, pre-test of evaluation matrix and guide and collection of evaluation reports .</p> <p>End of January 2005 in New York</p>		<p>March/April 2005</p> <p>Desk Assessment</p> <p>The team undertakes the quality assessment (QA) of the written reports. The results are analyzed among other things to identify potentially "good" or "instructive" practice as this relates to quality in UNFPA evaluations for follow up in the case studies.</p> <p>Total Estimated Time: up to max of 22 days for both Team Leader and Team Member. Evaluation Advisor estimated inputs up to 27 days.</p> <p>New York Interviews</p> <p>The team will conduct interviews in NY HQ with key UNFPA staff and compare on-going desk assessment results.</p> <p>Total Estimated Time: 4 days Team Leader; 3 days Team Member and 5 days Evaluation Manager</p> <p>Undertake Case Studies</p> <p>Seven instructive cases/examples are identified for in-depth follow up. The case studies are intended primarily as a means of getting at issues related to evaluation use and process issues.</p> <p>Total Estimated Time: 20 days Team Leader; 20 days for Team Member and Evaluation Advisor.</p> <p>Reference Group Meeting in New York</p> <p>Prior to beginning Phase 3, the Team will debrief the RG on key findings for testing out key recommendations. Other Programme Committee members could be included as needed.</p> <p>Team Leader and Team Member: 5 days each. Evaluation Advisor inputs as needed.</p> <p>Deadline: 17 June 2005 for aide memoire</p> <p>Late March/early April 2005</p> <p>Mid April. To Mid June 2005</p> <p>Late June 2005</p>	
February 2005			

PHASE	Key Products & Processes	PHASE	Key Products & Processes
3. Report Drafting		4. Report Validation, Dissemination & Finalization	
July 2005	<p>Production of First Draft of Full Evaluation Report</p> <p>Lead consultant produces first draft of evaluation report, which is not to exceed 40 pages (excluding annexes).</p> <p>Total Estimated Time: 20 days by Team Leader and 4 days by Team Member. Evaluation Advisor inputs as needed.</p> <p>Submission Deadline for first draft of report to DOS: 28 July 2005</p>	<p>August 2005</p> <p>Circulation of Draft Report to obtain written feedback and comments</p> <p>DOS circulates report widely within the organization for comment and reaction. (e.g. all CSTs, Country case study countries, Evalnet) E mail discussion group also formed.</p> <p>Mid September 2005</p> <p>Validation Workshop with Reference Group plus others as needed</p> <p>Workshop held at HQ for validating and disseminating EQA results. Special focus on developing priority actions for positive follow up and approval by Programme Committee. E mail discussion group used to further develop field inputs</p> <p>Final Report Submitted to DOS</p> <p>1 October 2005</p> <p>Final Presentation to Programme Committee</p> <p>DOS presents meta evaluation results and recommendations to PC for consideration and decisions on follow up actions.</p> <p>October 2005 meeting</p> <p>Total Estimated Time:</p> <p>Team Leader and Team Member 4 days each; Evaluation Advisor inputs as needed.</p>	

Annex 2

Sampling, assessment process, the EQA matrix and data analysis

A4.1 SAMPLING

To begin the sampling process, the evaluation team examined Country Office Annual Reports (COARs) covering the years 2002 through 2004. It was estimated that UNFPA Country Offices (and excluding HQ units) produce on average some 120 decentralized evaluations each year, or 360 reports for the three years covered by the EQA.

Given time and resource constraints, we deemed random sampling to be impractical. The size of a random sample required for a population of some 360 evaluations would have been relatively large, at 186.¹ With no direct access to evaluation reports from a central database or repository, the team would have faced substantial time delays in trying to get this number of reports assembled within the EQA timeframe.

The team decided to proceed with a smaller non-random sample, purposefully selected against a number of pre-determined criteria. A target of 80 evaluation reports from a representative cross section of countries and divisions was chosen. This was considered a reasonable number of evaluations, which the team could assess given available time and resources.

The team adopted a voluntary approach in anticipation of quickly receiving a significant number of evaluations, which UNFPA managers considered to be of good quality and whose results were being used in decision making. This approach may have introduced some bias into the sampling procedure by potentially including a higher number of what UNFPA managers perceived to be better quality evaluations.

Each division was provided with a target number of reports based on analysis of resource flows in 2002 and 2004. The results are provided in the Table below:

Division	Target number of evaluations	Number of evaluations received which met all the selection criteria (% of target)
Africa Division	29	22 (76)
Asia & Pacific Division	25	19 (76)
Latin America	7	7 (100)
Arab States & Europe	10	8 (80)
Interregional & Global	9	4 (44)
Total	80	60 (75)

¹ Assumes a 95 per cent confidence level and a +/- 5 per cent sample error.

Incoming evaluations were screened as to how well they met the selection criteria established by the meta-evaluation team. Evaluations not meeting the established selection criterion were excluded.

The following selection criteria were provided to Divisions:

Inclusion Criterion

- Examples of evaluations which managers found to be useful and used, either in the sense of having been influential or having made a positive difference.
- At least one or two examples of the following types of evaluations:
 - pilot project evaluations
 - Country Programme evaluations
 - Sub-programme or sub-component evaluations
 - thematic evaluations
 - joint evaluations
 - inter-regional and regional evaluations
- No more than three evaluations per country.
- Evaluations carried out after 2002.

Exclusion Criterion

- Any evaluation reports not managed by UNFPA
- Evaluation reports conducted before 2002

We also anticipated that there would be a reasonable spread across the different kinds of evaluation reports completed by UNFPA. The actual sample broke down as follows:

Type of evaluation	Number of evaluations (%)
Country Programme	9 (15)
Sub-programme	5 (8)
Project	28 (47)
Thematic	4 (6.5)
Regional	10 (17)
Sub-regional	4 (6.5)

Mid-term Review reports were included based on the requirements in the 1997 *Policies and Procedures Manual* section on Mid-Term Programme Review and the memo from the Deputy Executive Director (Programming) on Revised Programme Guidelines formats from April 2001, noting that Mid-term Review reports should include the nine evaluative criteria covered in the evaluation section of the 1997 *Policies and Procedures Manual*.

A4.2 DEVELOPMENT OF THE EQA MATRIX

The EQA Matrix is based on UNFPA results based management and evaluation guidance and tools, the particular needs of the UNFPA meta-evaluation, and general good practice. It drew heavily on the UNICEF meta-evaluation (2004) and the UNICEF *Evaluation Report Standards* (2004a), as well as ALNAP's (2004) meta-evaluation work, and reviews of the UN Evaluation Group and OECD-DAC draft evaluation standards. The *Programme Evaluation Standards* developed by the Joint Committee on Standards for Educational Evaluation, being the main standards used in the evaluation field, were a reference point throughout.

An initial matrix was developed and presented to the EQA Reference Group in January 2005. After this the meta-evaluation team spent three days editing the final draft of the matrix. The matrix was pre-tested on five evaluation reports. During the pre-test, four readers undertook a blind assessment of all five reports, making use of the draft EQA Matrix. Results were then compared and adjustments made to individual standards in cases where there had not been a high level of consistency achieved among meta-evaluators in assigning ratings. Some more minor adjustments were made later on during the data collection phase as and when needed. For bias and limitations, see Annex 7.

A4.3 RATING SCALE

Deciding on a rating scale took a considerable amount of time – unsurprisingly, as this is one of the key features of a meta-evaluation assessment. There were a number of factors taken into account, in particular the cut off point for satisfactory quality, the fact that 'odd' rating scales may lead to pooling around the middle rating, and the rating scale balance (i.e. how many negative and positive elements to include). All rating scales have their drawbacks and advantages, and it was difficult for the team to come to a final decision. A five point rating scale was finally selected:

- Missing: the standard is not addressed
- Unsatisfactory: the standard is addressed but not at the level specified
- Satisfactory: the standard is met
- Good: the standard is exceeded
- Excellent: Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports

A further category of 'not applicable' was added where the standard was viewed as not relevant for a particular evaluation. The main area where this category was applied was for the DAC criteria of 'sustainability' and 'impact' in Mid-term Reviews; however, the not applicable category was applied in only 20 cases (0.006 per cent of total ratings). Clear definitions were developed for each of the 28 standards in the matrix so as to ensure consistent rating by meta-evaluators. See the EQA matrix below for further details where the definitions for each standard are included.

A4.4 ASSESSMENT PROCESS

Two meta-evaluators read and rated each evaluation independently, and then compared ratings. It was important to have two meta-evaluators reading each report as this reduced bias and error. It also fed into discussions of evaluation quality which was useful for the good practice country case studies. During the comparison of rating obvious errors were changed, and for areas of continued disagreement meta-evaluators returned to the evaluation and made changes to their rating where



they thought this was relevant. The 'rationale for rating column' was important for assuring that the four meta-evaluators reached their ratings from similar perspectives, and as a reminder to meta-evaluators as to why they gave a particular rating.

Shadow rating worked well. Rating achieved a consistency of 91 per cent between the meta-evaluators, which is in-line with other similar exercises (e.g. ALNAP 2004). The main issues were:

- The difference between 'missing' and 'unsatisfactory'. For example, one or two passing references to gender equality in an evaluation could be rated in either of these categories. For the sake of convenience it was decided that even if an area was mentioned only once in passing it would be rated 'unsatisfactory' rather than 'missing'. This has likely led to a more positive estimate of evaluation quality, but is not seen as biasing overall results since the main cut-off in analysis is between 'satisfactory' and 'unsatisfactory'.
- Rating 'excellence'. As the team decided that this rating would be a comparison to the sample as a whole, meta-evaluators had to return to matrices and change ratings to reflect good practices they had indicated earlier.
- Individual criteria. In particular, the evaluation's contribution to policy dialogue, and the difference between evaluation objectives and purpose, and the link between the objectives and methodology, were sometimes difficult to rate.

A4.5 DATA ANALYSIS

A rating was considered a match between the two meta-evaluators if both ratings were the same, or above the 'satisfactory' cut-off point (i.e. satisfactory-good, good-excellent, and satisfactory-excellent were considered matches; missing and any other rating, and unsatisfactory-satisfactory, unsatisfactory-good, and unsatisfactory-excellent were not considered matches). Where one meta-evaluator rated higher than the other but there was still a match, the average rating was taken during analysis (e.g. satisfactory was given a score of '2' and good a score of '3', and where there was one rating on each of these the overall score was 2.5). Non-matched and non-applicable ratings were not included in the analysis.

Data were fed into an Excel spreadsheet, and were cross-checked for entry and calculation error. After this the incidence of ratings was added to give an overall rating, by individual criteria, by region, and for the sample as a whole.

Ratings were also aggregated for individual evaluations to facilitate intra-sample comparisons (for example performance of different kinds of evaluation types) and testing of hypotheses (e.g. whether higher quality terms of reference lead to higher quality evaluations). No weighting was assigned to individual criteria for this aggregation; ratings for both meta-evaluators were added, with missing scored as '0', unsatisfactory as '1', satisfactory as '2', good as '3' and excellent as '4'. The aggregate figure was divided by the number of criteria in each case (having removed non-match and non-applicable ratings) to give the aggregate rating for individual evaluations.

GUIDE TO UNFPA META-EVALUATION MATRIX

April 5, 2005

TOPICS	
TOR	Guide
1.a. Purpose/Rationale of the evaluation is clearly defined.	<p>WHAT is the standard The purpose/rationale explains why the intervention is being evaluated.</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Does not adequately explain why the intervention is being evaluated. • Satisfactory. Provides adequate detail as to why the intervention is being evaluated. • Good. Provides a full account as to why the intervention is being evaluated. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
1.b. Use and users clearly defined.	<p>WHAT is the standard The evaluation clearly identifies what will be the use of the evaluation, its users and target audience.</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. The use, users and target audience are unclear. • Satisfactory. Main users and use are detailed. • Good. Main users and use are detailed, along with strategies for follow-up to recommendations and/or lessons learned. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
1.c. Scope and focus is clearly identified	<p>WHAT is the standard The terms of reference need to set out clearly the evaluation's intended scope and focus - that is the main areas which the evaluation will cover and its main topics of interest. 1) The ToR should either define the main evaluation questions, and/or list objectives. 2) Objectives should relate to the purpose and be precisely stated so they guide the evaluator in terms of information needs and data to collect. Evaluation questions should be realistic and achievable. 3) The ToR should specify evaluation criteria to be used given the evaluation's objectives and scope, such as the OECD/DAC criteria (relevance, efficiency, effectiveness, sustainability and impact).</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. The evaluation covers 1 and 2 above only partially. • Satisfactory. The evaluation covers points 1 and 2 above. • Good. The evaluation fully explains the objectives and/or evaluation questions and presents the key criteria (such as the DAC criteria) providing detail on each of the criteria. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
1.d. Methodology specified	<p>WHAT is the standard</p> <p>The ToR should define: 1) information sources for data collection; 2) sampling procedures, including area and population and sample size, 3) data collection instruments (e.g. review of literature, focus groups, semi-structured interviewing, survey); 4) data analysis methods; 5) measures expected to ensure that the evaluation process is ethical and that confidentiality and dignity of participants in the evaluation -e.g. interviewees, sources - will be protected</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. The evaluation only partially describes the methodology, e.g. one or all of points 1-5 are partly covered, or two points are adequately covered. • Satisfactory. Adequately covers points 1-3. • Good. Adequately covers points 1-4. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
1.e. Roles and responsibilities of evaluators defined	<p>WHAT is the standard</p> <p>The TOR should specify the roles and responsibilities of the evaluation team leader and team members, as well as other stakeholders and advisory structures involved, e.g. steering committees. It should clarify who is responsible for: 1) liaison with the evaluation team; 2) providing technical guidance; 3) coordinating the stakeholders involved; 4) selection, orientation and training of team members, and data collection assistants, where applicable.</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Only partially describes the roles and responsibilities. • Satisfactory. Adequately covers at least two of points 1-4. • Good. Adequately covers points 1-4. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
1.f. Clear outputs, deadlines, formats specified	<p>WHAT is the standard</p> <p>The ToR should explain clearly the outputs and/ or products to be produced by the evaluation. The following should be included 1) the evaluation report; 2) methodology, data sets, and list of interviewees; 3) dissemination material (e.g. two page summaries, presentation materials); 4) formats for outputs/products, including software, number of hard copies, translation requirements, structure and length of the evaluation report, and deadlines.</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Partly covers one or more of the points above. • Satisfactory. Partly covers all or fully covers two of the points above. • Good. Provides a full account of the outputs, deadlines and formats as per the description above. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
The Evaluation Report	
2. Completeness of report	<p>HOW is the standard applied</p> <p>Presence of:</p> <ul style="list-style-type: none"> • Executive Summary: Y or N • ToR: Y or N • Bibliography: Y or N • Data collection instruments: Y or N • List of people met (name, title, location): Y or N
3. Quality of Executive Summary	<p>WHAT is the standard</p> <p>An Executive Summary should provide an overview of the essential parts of a report. It should be short, concise, readable and well organized – and should “stand alone” (without requiring reference to the rest of the report.) The Executive Summary should include:</p> <ul style="list-style-type: none"> • Brief description of the programme/project • Context of the programme/project – years of implementation, situation vis-à-vis UNFPA Country Programme outcomes and other programming it contributes to (i.e. UNDAF outcomes, complementary national or partner programmes) • Basic description of context and purpose of evaluation – why this evaluation now • Objectives of evaluation • Key features and methodology • Most important findings and conclusions • Key recommendations <p>How is the standard applied:</p> <ul style="list-style-type: none"> • Missing – the report lacks an Executive Summary. • Unsatisfactory – the Executive Summary does not provide the whole picture, leaving out essential information, such as key findings, conclusions and/or recommendations. • Satisfactory – the Executive Summary can stand alone and includes key recommendations, conclusions and findings, but is missing other elements above. • Good – the Executive Summary includes all of the elements above and can stand-alone • Excellent – Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
4. Purpose of the evaluation is outlined	<p>WHAT is the standard</p> <p>The evaluation should outline why it is being carried out. Purpose should not be confused with the evaluation objectives, which state what the evaluation seeks to accomplish. The purpose also relates to the timing of the evaluation in the project cycle – at the beginning of the project, mid-way through the project, or after the project is complete. Sometimes it may be relevant to relate the programme/project to the Country Programme cycle, especially if the evaluation is to contribute to a Mid-Term Review.</p> <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory – Purpose is unclear. • Satisfactory – Purpose is clear, with specific justification for timing of the evaluation. • Good – Purpose is clear and relates appropriately to objectives, and specific justification for timing of the evaluation is given. • Excellent – Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
5. Use and users clearly defined	<p>What is the standard?</p> <p>How will the evaluation be used? Who will use it? Explaining why the evaluation is being done - what triggered the evaluation - does not always answer how it will be used by stakeholders. For example, an evaluation may state that it is being carried out at the end of a five-year programme. The reader is left wondering if the evaluation is to assist the programme in the next cycle, or to determine how to hand the programme over to the government, an NGO, or for general lessons learned for similar programmes. Evaluations are most used when they are planned to coincide, or are driven by a decision that needs to be taken, or when evaluators work with intended users to define methods and recommendations. Ideally, an evaluation is also intended to be of use to partners and other stakeholders.</p> <p>How is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory- Uses and users not clearly defined. • Satisfactory- Main use and users are explicitly set out. • Good- Main use and users of evaluation are explicitly set out, and this is linked to follow-up strategies and recommendations. • Excellent- Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
6. Objectives of evaluation clearly defined	<p>WHAT is the standard</p> <p>Evaluation objectives should provide a clear statement of what the evaluation seeks to accomplish. Objectives can be detailed further in specific evaluation questions. The evaluation should also demonstrate how the objectives follow from the purpose.</p> <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - objectives are vague, and do not provide a clear statement of what the evaluation seeks to accomplish. • Satisfactory - objectives provide a clear statement of what the evaluation seeks to accomplish. • Good - objectives provide a clear statement of what the evaluation seeks to accomplish. Evaluation questions are detailed. Objectives are linked to the evaluation purpose, and/or are linked to the OECD-DAC criteria (efficiency, effectiveness, relevance, sustainability and impact). • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	TOR	Guide
7. Quality of methodology		<p>WHAT is the standard</p> <p>The methodology adopted should follow good practice and be appropriate. The methodology should be clearly outlined in the evaluation report.</p> <p>The standard covers three areas:</p> <p>7a.Logical link to evaluation objectives and evaluation questions</p> <ul style="list-style-type: none"> • The methodology adopted should be valid, that is it should focus on the evaluation objectives and questions. <p>7b.Methodology follows good practice</p> <ul style="list-style-type: none"> • Data should come from a variety of sources to ensure its accuracy, and also to ensure that all affected people/stakeholders are considered. Data should be triangulated (i.e. cross-checking of quantitative and qualitative data and different types of data sources i.e. interview, observation, and document analysis). • Description of data collection methods and analysis should be included in the evaluation report (including methods for analysis of quantitative and qualitative data). • Description of sampling should be included - area and population to be represented, rational for selection, mechanics of selection, numbers selected out of potential subjects, and limitations of sample. • The methodology should explicitly address issues of gender and marginalized groups. • Commonly accepted practice for the given situation by evaluation professionals may also be considered and referenced. <p>The appendices should include the following in addition to more detail on any of the above:</p> <ul style="list-style-type: none"> • Data collection instruments (surveys, checklists, etc.). <p>7c.Efforts to control bias and acknowledgement of limitations</p> <ul style="list-style-type: none"> • The evaluation should set out clearly any limitations in the methodology employed. Limitations can come from a variety of sources both internal and external. Bias can be at three levels: • Sources of data - the respondents themselves have a bias in their opinion on the topic. • Methods of data collection - the structure of the data gathering could be skewed to favour one factor, preconceived idea or viewpoint, or miss key areas. • Analysis of data - the evaluators have a bias towards a certain viewpoint that colors their interpretation of the findings. <p>HOW is this standard applied - nb rate separately for each area 7a, 7b and 7c</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - the area is only partly covered. • Satisfactory - the area is covered but with limited detail. • Good - the area is well covered with substantial detail. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
8. Considerations given to propriety and ethics	<p>WHAT is the standard</p> <p>The evaluation report should contain a description of the measures and mechanisms put in place to: ensure that the evaluation process was ethical, that stakeholders were protected, and address any ethical dilemmas or issues that emerged.</p> <p>The design should include the following areas:</p> <ul style="list-style-type: none"> • the balance of costs and benefits to participants including potential negative impact, • the ethics of who is included and excluded in the evaluation and how this is done, • handling of privacy and confidentiality, • practices of obtaining informed consent, and • feedback to participants. <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - scant attention to ethical considerations (e.g. one or two sentences). • Satisfactory -description of ethical considerations included addressing the areas above. • Good - good description of ethical considerations fully addressing the areas above. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
9. Evaluator values and bias outlined	<p>WHAT is the standard</p> <p>The following areas should be covered:</p> <ul style="list-style-type: none"> • The perspectives, procedures and rationale used to interpret the findings should be carefully described, so that the bases for value judgements are clear. • Political and other beliefs and how these may have influenced evaluation results should be made clear. <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Incomplete statement concerning values and potential bias. • Satisfactory. Both of the areas above are covered. • Good. Both of the areas above are covered in detail. • Excellent. Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
10. The project/programme to be evaluated was clearly described	<p>WHAT is the standard</p> <p>The program being evaluated should be described and documented clearly and accurately, so that the project/programme is clearly identified.</p> <p>The following areas should be covered:</p> <ul style="list-style-type: none"> • The ways in which national and/or sub-national context shaped project/ programme design need to be outlined. • Sufficient details about the project/programme should be included so that the reader can assess whether findings, conclusions and recommendations are valid (e.g. size of project/programme; numbers of people intended to be reached, broken down by sex; and budget figures). • Sufficient details should be included as to how the project/programme was intended to address problems identified, including a causality analysis.⁸ • The overall goal that the programme/project was expected to contribute to and how it was expected to contribute should be described. The overall goal may be related to national development plans and/or MDGs, which may in turn be supported by UNFPA goals. • Any changes in project/programme design and implementation should be outlined. <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - a vague programme/project description is given (e.g. project/programme logic is not described, lack of attention to context, lack of causality analysis) • Satisfactory - all areas above are covered • Good - all areas above are covered in substantial detail • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
11. The role and contributions of UNFPA and other stakeholders were clearly defined	<p>WHAT is the standard</p> <p>The reader needs to have a sense of who is doing what to facilitate an understanding of which partners contributed to the results, taking into consideration that attribution of specific results in joint programming is often problematic. The evaluation should outline who is involved, in what role, and what they have contributed to the programme/project including: financial resources, in-kind contributions (material items such as drugs, books, desks, etc.), technical assistance, participation, staff time, training, leadership, advocacy and lobbying. This should include any contributions from primary stakeholders</p> <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - stakeholder contributions are inadequately outlined (e.g. only a single party's contributions are set out). • Satisfactory - all stakeholders' contributions are outlined. • Good - all stakeholders' contributions are outlined in a way that facilitates an understanding of which partners contributed to which results. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

8. Causality Analysis: A type of analysis used in programme formulation to identify the root causes of development challenges. Development problems often derive from the same root causes (s). The analysis organizes the main data, trends and findings into relationships of cause and effect. It identifies root causes and their linkages as well as the differentiated impact of the selected development challenges. Generally, for reproductive health and population problems, a range of causes can be identified that are interrelated. A "causality framework or causality tree analysis" (sometimes referred to as "problem tree") can be used as a tool to cluster contributing causes and examine the linkages among them and their various determinants. (UNFPA glossary)



TOPICS	
TOR	Guide
12.Extent of Stakeholders/ beneficiaries involvement	<p>WHAT is the standard</p> <p>The degree of participation of stakeholders in the evaluation process can vary along a continuum from low to high based on what key steps or activities stakeholders are involved in - some steps are more pivotal than others in shaping results - and what role stakeholders can have in each. While not all evaluations can be participatory to the same degree, it is important that consideration is given to participation of stakeholders, and that the evaluation report is transparent about the rationale and level of participation of different stakeholders.</p> <p>Roles of stakeholders include:</p> <ul style="list-style-type: none"> • Identifying the purpose • Identifying the objectives • Formulating the questions • Designing the methodology • Designing data collection instruments • Collecting data • Analysing data • Writing the report • Establishing a follow-up action plan <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing - no description of stakeholder participation is given (although stakeholders may be involved in the evaluation as data sources). • Unsatisfactory - vague description of stakeholder participation is given, and/or a very limited definition of "participation" or "stakeholder" is used, e.g. no attention to primary stakeholders. • Satisfactory - clear description of stakeholder participation but no rationale provided for the degree of participation decided upon. • Good- clear description and justification of stakeholder participation given, including reference to primary stakeholders. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports. <p>LIMITATIONS to the use of this standard</p> <p>The breadth and degree of stakeholder participation feasible in evaluation activities will depend in part on the kind of participation achieved in the programme/project. Nonetheless, evaluation activities can be used to promote greater participation. Ideally there will be a few strategically important evaluation activities where a broader range of stakeholders can be brought together to explore common research/evaluation questions.</p>

TOPICS	
TOR	Guide
13. Quality of assessment of gender mainstreaming	<p>WHAT is the standard</p> <p>'Mainstreaming a gender perspective is the process of assessing the implications for women and men of any planned action, including legislation, policies or programmes, in all areas and at all levels. It is a strategy for making woman's as well as men's concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of policies and programmes in all political, economic and societal spheres so that women and men benefit equally and inequality is not perpetuated. The ultimate goal is to achieve gender equality'. (ECOSOC Agreed Conclusions)</p> <p>The evaluation should assess the extent to which:</p> <ul style="list-style-type: none"> • The intervention effectively used sex-disaggregated data. • The intervention promoted gender mainstreaming and improved gender relations. <p>(Source: based on Gender, Culture, Human Rights Branch, TSD - Intranet Site.)</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - gender and/or women mentioned but above standards inadequately addressed. • Satisfactory - both bulleted areas are addressed. • Good - both bulleted areas are addressed in detail (i.e. there is detailed discussion of both of the points above, with some reference to the effects of the intervention on gender equality). • Excellent. Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
14. Quality of assessment of capacity-development	<p>WHAT is the standard</p> <p>Capacity is the ability of individuals, organizations and systems, including networks of organizations, to perform in support of their development objectives. Success in capacity development comes from addressing it from a systems perspective and identifying the type of interventions - resources, techniques, knowledge - at the individual, organizational and systems levels that can make a genuine difference to overall systems performance. Capacity is developed in the course of an iterative and long-term process of learning and adaptation. It is developed by the participants in the process "while doing". Capacity is only developed if participants in the process support and believe in what they create [i.e. local ownership] (Source: Thematic Evaluation Report #20: UNFPA's Support to National Capacity Development - Achievements and Challenges. UNFPA. 2003.)</p> <p>The evaluation should assess the extent to which the intervention applied the following capacity development strategies:</p> <ul style="list-style-type: none"> • Drew upon local cultural characteristics and systems ("the local way of doing business") in designing and implementing the intervention; • Provided relevant training; • Developed a knowledge base and promoted its use; • Developed systems, including planning, monitoring and evaluation systems, to improve performance; • Strengthened and promoted partnerships and networking; • Developed a common vision of internationally approved objectives in Reproductive Health, and population and development through advocacy and policy dialogue. <p>(Source: Adapted from the UNFPA MYFF Strategies, 2004-2007.)</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - addressed only one capacity development strategy (for example training). • Satisfactory - addressed bullet point one and one other capacity development strategy. • Good - adequately addressed all of the capacity development strategies while assessing the ways in which these interacted to create sustainable capacities and improved performance • Excellent. - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
15. Quality of assessment of human rights	<p>WHAT is the standard</p> <p>A Statement of Common Understanding was developed in the Interagency Workshop on a Human Rights Based Approach (HRBA) in the context of UN reform 3-5 May, 2003 that identifies the necessary elements for a HRBA; the application of “good programming practices” does not by itself constitute a human rights-based approach, and requires additional elements.</p> <p>The following elements are necessary, specific, and unique to a human rights-based approach:</p> <ol style="list-style-type: none"> Assessment and analysis identify the human rights claims of rights-holders and the corresponding human rights obligations of duty-bearers as well as the immediate, underlying, and structural causes of the non-realization of rights. Programmes assess the capacity of rights-holders to claim their rights, rights and of duty-bearers to fulfil their obligations. They then develop strategies to build these capacities. Programmes monitor and evaluate both outcomes and processes guided by human rights standards and principles. Programming is informed by the recommendations of international human rights bodies and mechanisms. In programming, UNFPA must be mindful of the articles of the ICPD Programme of Action and CEDAW and of the guiding principles of these and other human rights treaties. <p>Human rights principles particularly relevant to the ICPD and how to meet them include:</p> <ul style="list-style-type: none"> Right to life, survival and bodily integrity - for instance by preventing avoidable maternal deaths; ending female feticide and infanticide; Right to liberty and security of the person - for instance by eliminating female genital mutilation; obtaining informed consent for all procedures, including HIV testing, sterilization, and abortion; Right to freedom from inhuman and degrading treatment - for instance by prohibiting involuntary abortion and sterilization; protecting and caring for survivors of sexual assault and domestic abuse and prosecuting the perpetrators; Right to marry and found a family - for instance by preventing early and coerced marriages; Right to decide the number and spacing of one’s children - for instance by providing access to a range of contraceptive methods, and providing access to safe abortion services, where legal; Right to the highest attainable standard of health - for instance by providing access to affordable, acceptable and comprehensive health services, including reproductive and sexual health services for women and men; Right to the benefits of scientific progress - for instance by providing access to emergency contraception or to obstetric care that can prevent maternal deaths; Right to non-discrimination and respect for diversity for example, the rights of all to reproductive health services including adolescents, unmarried women, migrants, refugees; Right to receive and impart information - by making family planning information and services freely available; Right to freedom of thought, conscience and religion - for instance by not limiting RH services on religious grounds Right to privacy - for instance by ensuring privacy for all services Right to development - particularly the rights of the poor to a decent standard of living and freedom from want Right to freedom of movement - such as protecting migrants’ rights; Right of everyone to education, enabling all persons to participate effectively in society. “ <p>(Adapted from UNFPA Knowledge Asset on Human Rights - asset under construction and OUTLOOK, Volume 20 Number 4. December 2003.)</p>

TOPICS	
TOR	Guide
15. Quality of assessment of human rights (continued)	<p>The evaluation should outline and analyse how the intervention:</p> <ul style="list-style-type: none"> • was informed by the ICPD Programme of Action as listed above, and CEDAW; • identified the human rights claims and obligations relevant to the project/programme; • Identified gaps in the capacity of rights-holders to claim their rights, and of duty-bearers to fulfil their obligations, including (but not solely) an analysis of gender and marginalized and vulnerable groups; • addresses identified gaps; and • monitored results. <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - minimal consideration given to this approach (e.g. a passing or token reference). • Satisfactory - the evaluation meets the first three areas in the bullet points immediately above, but not in an in-depth manner. • Good - the evaluation provides an in-depth analysis that covers all five areas in the bullet points immediately above. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
16. Evaluation enables UNFPA to engage in evidence based policy dialogue	<p>WHAT is the standard</p> <p>UNFPA funded evaluations should contribute evidence-based arguments, which can be used to support advocacy efforts in the broader policy arena. In order to do so, evaluations must be designed and used with this goal in mind. Evaluations therefore must to the extent possible meet the following areas:</p> <ul style="list-style-type: none"> • Use quantitative and qualitative data to analyse programme/project implementation processes, outputs, outcomes and impact as well as contextual factors that facilitated or constrained the achievement of results; • Use causality and institutional and organizational context analysis to explain outcomes and impact; • Analyse whether the programme/project intervention was based on an in-depth understanding of the vision and interests of the objects for policy dialogue (i.e. stakeholder analysis); • Analyse whether the programme/project intervention promoted active partnerships and alliance building; • Use evaluation analysis to document new approaches and successful interventions which can be brought to scale by government or other development partners; and • Interpret evaluation findings in relation to larger policy debates in the particular programme/project context or related to the ICPD PoA, MDGs and provide related recommendations for UNFPA policy dialogue and advocacy. <p>HOW this standard is applied:</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - evaluation does not interpret evaluation findings in relation to national or international policy debates. • Satisfactory- evaluation uses quantitative and qualitative data to analyse intervention processes leading to results highlighting constraining and facilitating factors. Evaluation places this analysis in the context of policy debates at national and international levels (relating to International conference recommendations such as ICPD PoA, MDGs) but does not provide operational policy recommendations for UNFPA. • Good- evaluation satisfies in-depth all the criteria of “satisfactory” and provides operational policy and advocacy recommendations for UNFPA. • Excellent- - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
17. Quality of assessment of cultural issues	<p>WHAT is the standard</p> <p>Culture is the whole complex of distinctive spiritual, material, intellectual and emotional features that characterize a society or a group. It includes creative expressions, community practices and material or built forms” UN World Commission on Culture and Development. 1995. Culture is learned during the process of socialisation in formal and informal settings which include family, school, religion and employment; it is not inherent nor ascribed. Altogether culture sets our standards, tastes, etiquette and morality, values, meanings and attitudes to everything. Culture is also dynamic and it can thus change through exposure to new knowledge, science and technology, or as people acquire new needs and goals.</p> <p>(Source: UNFPA’s Knowledge Asset on Mainstreaming Culturally Sensitive Approaches in UNFPA’s Programming).</p> <p>Based on the above definition of culture and UNFPA’s approach in this area, the evaluation should, as an integral part of its findings, conclusions and recommendations, discuss how the design and implementation strategies of the intervention took into account the cultural characteristics of the society and stakeholders.</p> <p>In particular, the evaluation should assess whether cultural aspects were adequately considered in the design and implementation of the intervention, through:</p> <ul style="list-style-type: none"> • avoiding value judgments by grounding project design and implementation on evidence. • talking to people, communities, community leaders, organizations, involving them in intervention design and implementation; • basing interventions on local experience, skills and vision of what local communities and organizations defined as their own physical, psychological and material well-being; • identifying, exploring and understanding local support systems, structures, norms and values that can be used and developed to promote universal human rights including the right to reproductive health, gender equity and equality as well as national ownership of the ICPD agenda <p>HOW the standard is applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - culture is mentioned as being a factor in the design and implementation of the intervention without explaining how. • Satisfactory - the evaluation addresses the first three areas in the bullet points immediately above • Good - the evaluation addresses all four areas in the bullet points immediately above; and provides conclusions and recommendations that address the findings related to cultural issues. • Excellent. Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports

TOPICS	
TOR	Guide
<p>18. Quality of assessment of national ownership</p>	<p>WHAT is the standard</p> <p>Local or national ownership can be defined in many ways. The following is a good definition by Van de Walle, N. And T. Johnston, eds. (1996). <i>Improving Aid to Africa</i>. Baltimore: John Hopkins: "Recipient governments and beneficiaries can be said to "own" an aid activity when they believe that it empowers them and serves their interests... From a sense of ownership flows a willingness to commit real resources to ensure the activity's success, to solve problems that emerge during implementation and to sustain the activity after the withdrawal of aid".</p> <p>(Source: UNFPA TSD Intranet Site on SWAps.)</p> <p>Country Programmes should, to the extent feasible, build on existing national, non-governmental and civil society systems and processes in order to enhance long-term capacity development. Government leadership is a prerequisite for the success of UNFPA supported programmes.</p> <p>Source: as quoted in UNFPA PPM.</p> <p>The evaluation should analyse the extent to which the intervention:</p> <ul style="list-style-type: none"> • is consistent with partners' priorities and effective demand; • is supported by local institutions and well integrated with local social and cultural conditions; • involved national stakeholder participation in planning, implementation and M&E; • involved government or community co-financing or development of a system to ensure future local flow of funds to the intervention. <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Addresses one of the above areas in a superficial manner. • Satisfactory. Addresses three or more of the above areas, but without in-depth analysis • Good. Detailed analysis of all of the above areas. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
19. Quality of assessment of partnership and alliance building	<p>WHAT is the standard</p> <p>Partnerships comprise an inter-organizational strategic collaboration to achieve shared short and/or long-term results in which the efforts of each partner are coordinated with and enhance the efforts of others. It encompasses all stakeholders whose contribution could help achieve an outcome or output. The new UNFPA Strategic Direction and Multi-Year Funding Framework (MYFF) require a strong emphasis on multi-sectoral partnerships.</p> <p>Partnerships with other donors and between government agencies increasingly occur through policy-level committees that develop or oversee PRSPs, SWAps, MDGs, and national development plans. Through these partnerships, UNFPA's main objective is to achieve incorporation of RH and gender and explicit linkages of population and poverty in these plans. An equally important focus of the MYFF is building, strengthening, and promotion of partnerships among state and civil society actors.</p> <p>There are different levels of "strategic alliances" according to the degree of integration that partners aim to achieve: Networking: is an information exchange mechanism in which members undertake no or very few joint tasks. Cooperation: members identify mutual needs, and undertake joint strategies and tasks. Partnering: organizations share resources and develop strategies to reach mutual goals, with a more central body of people representing their organizations, and formal communications and decision-making mechanisms. Merging: organizations merge resources to create a new entity or consortium that undertakes a long-term commitment to achieve short and long-term outcomes. Has a formal structure, committees, and well-defined roles and responsibilities. The decision on level of integration depends on which level best serves the desired output.</p> <p><i>Source: Powerpoint presentation on MYFF strategies. Strategic Planning Office. July 2004.</i></p> <p>Based on the above definition of partnerships, the evaluation should assess:</p> <ul style="list-style-type: none"> • how the intervention addressed the development/strengthening of partnerships and alliances to achieve planned intervention results • whether the intervention was based on a stakeholder analysis that identified the different levels of possible strategic alliances to achieve planned intervention results • whether specific results and indicators for partnerships building were identified <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - partnership and alliance building is mentioned either as a present or absent dimension of the intervention without further elaboration. • Satisfactory - partnership and alliance building is described either as a present or absent dimension of the intervention. The evaluation addresses the three standards described above. • Good - the evaluation provides an in-depth analysis of partnership and alliance building. It addresses all of the standards above, explaining them within the particular context of the intervention, and draws and makes related conclusions and recommendations. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
20. Quality of assessment of RBM	<p>WHAT is the standard</p> <p>Results Based Management (RBM) is a management strategy by which an organization ensures that its processes, products and services contribute to the achievement of desired results (outputs, outcomes & impacts). RBM rests on stakeholder participation and on clearly defined accountability for results. It also requires monitoring of progress towards results and reporting on performance/feedback which is carefully reviewed and used to further improve the design or implementation of the programme (UNFPA PM&E Glossary).</p> <p>The evaluation should explore underlying questions as to how a programme/project is managed and what information stakeholders have had access to and used to take key decisions in design and implementation. The evaluation should address the following dimensions of RBM:</p> <ul style="list-style-type: none"> • Existence of a well defined and results-oriented programme/project planning, monitoring and evaluation system (people, tools and processes) • Availability of indicators that the programme/project had planned to use to monitor programme/project performance • Indicators that programme/project implementers and decision-makers actually used • How programme/project stakeholders (beneficiaries, implementers, managers and other decision-makers) tracked the indicators and used them as well as other monitoring and evaluation mechanisms to assess programme/project performance and results and to adjust the programme/project objectives and strategies. <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory - minimal consideration given to RBM (e.g. a token paragraph on monitoring and evaluation). • Satisfactory - analyses bullets 1-3 above • Good - analyses in some detail all four bullets above. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
21. Quality of assessment with regards to relevance	<p>WHAT is the standard</p> <p>An assessment of programme/project relevance examines the degree to which the outputs, outcomes/ purpose and/or goals remain pertinent as originally planned or subsequently modified (cf. UNFPA M&E Glossary). Analysis should cover:</p> <ul style="list-style-type: none"> • whether the programme/project design is in line with national needs, policies and priorities of programme/project target groups; • whether the programme/project is in line with UNFPA's policies and priorities, especially the 1994 ICPD PoA; • synergy between UNFPA's intervention and that of other development partners; • whether programme/project results are relevant to stakeholders (e.g. have the right kinds of resources, training or information been provided). <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. The analysis meets only the first bullet in a summarized way. • Satisfactory. The evaluation provides an identifiable assessment of bullets 1, 3 and 4. • Good. All bullet points above addressed in-depth. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
22. Quality of assessment with regards to effectiveness	<p>WHAT is the standard</p> <p>Effectiveness is a measure of the extent to which a programme achieves its planned results (outputs, outcomes and goals) (UNFPA glossary).</p> <p>The evaluation should include the following:</p> <ul style="list-style-type: none"> • Causality analysis to explain how inputs and activities led to outputs, outcomes and impact. If it is a formative evaluation, it should assess whether inputs and activities are likely to lead to the planned outputs, outcomes and impact; • Assessment of coverage (e.g. was the planned areas and target group successfully covered?); • Assessment of constraining and facilitating factors and the influence of context on the achievement of results. <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Bullet points above inadequately covered (e.g. the evaluation assesses inputs only and/or does not use causality analysis) ; • Satisfactory. The evaluation provides an assessment of all of the bullet points above. • Good. The evaluation provides a detailed assessment of all of the bullet points above. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
23. Quality of assessment with regards to efficiency	<p>WHAT is the standard</p> <p>Efficiency is a measure of how economically or optimally inputs (financial, human, technical and material resources) are used to produce outputs (UNFPA glossary).</p> <p>The evaluation should include the following:</p> <ul style="list-style-type: none"> • assessment of the quality of outputs achieved in relation to the expenditures incurred, and resources used; • assessment of timeliness of inputs including personnel, consultants, travel, training, equipment and misc. costs and the timeliness of outputs; • whether there was adequate justification for the expenditures incurred and whether the resources were spent as economically as possible, taking into account possible alternatives. <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory. Discusses only inputs and does not relate this to achievement of outputs. • Satisfactory. Assesses quantity, quality and timeliness of inputs and links this to achievement of outputs. • Good. Assesses quantity, quality and timeliness of inputs and links this to achievement of outputs; and analyses whether the resources were spent as economically as possible, and potential alternative approaches that might have been more efficient. • Excellent. Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
24. Quality of assessment with regards to sustainability	<p>WHAT is the standard.</p> <p>Sustainability can be defined as the durability of programme results after the termination of the technical cooperation channelled through the programme. Static sustainability - the continuous flow of the same benefits, set in motion by the completed programme, to the same target groups; dynamic sustainability - the use or adaptation of programme results to a different context or changing environment by the original target groups and/or other groups. (UNFPA glossary)</p> <p>The evaluation should include the following:</p> <ul style="list-style-type: none"> • assessment of the extent to which the programme/project results have had or are likely to have lasting results after programme/project termination and the withdrawal of external resources; • assessment of the factors affecting sustainability on the basis of the priority assigned to the programme/project by stakeholders (e.g.. their readiness to continue supporting or carrying out specific activities; replicate the activities in other regions or sectors of the country); or adapting programme/project results in other contexts • assessment of the availability of local management, financial and human resources needed to maintain the programme/project results over the long term. <p>HOW is the standard applied.</p> <ul style="list-style-type: none"> • Missing. • Unsatisfactory. An evaluation that only states it will be sustainable but does not tell us why, and only partly meets the first two areas above. • Satisfactory. The likelihood of continued stakeholder support for or adaptation of the programme/project is explained and the evaluation adequately meets the first two areas above. • Good. Includes analysis of local capacity to maintain or adapt programme/project results and adequately addresses all three areas above. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
25. Quality of assessment with regards to impact	<p>WHAT is the standard</p> <p>Impact is the positive and negative long-term effects on identifiable population groups produced by a development intervention, directly or indirectly, intended or unintended. These effects can be economic, socio-cultural, institutional, environmental, technological or other types. (UNFPA Glossary)</p> <p>HOW is the standard applied</p> <ul style="list-style-type: none"> • Missing. May be not relevant because it is a mid term, interim, or process type evaluation. • Unsatisfactory. Does not attempt to analyze the intervention from the perspective of long-term change, or comment on whether long-term results are likely to be achieved. • Satisfactory. Gives appropriate and visible consideration as to whether long term results have been achieved or are likely to be met. Assessment does not need to be highly rigorous, nor does it need to make a lot of distinction among population groups or the different types of effects listed above. • Good. Meets satisfactory rating, and also applies causality analysis, and provides a better level of analysis about different types of impact. • Excellent. Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
26. Quality of conclusions	<p>WHAT is the standard</p> <p>Conclusions should add value to the findings. Conclusions should flow logically from, and reflect, the report's central findings. The evaluation report should cite the information that supports each conclusion, and provide a clear and defensible basis for value judgements and the allotment of praise or blame. This information should allow the reader to either accept or reject the conclusions and reflections of the evaluators.</p> <p>Conclusions must also focus on issues of significance to a programme/project. This choice of significant issues must relate back to the evaluation objectives and key questions.</p> <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory –Conclusions that do not follow from the findings presented; and/or conclusions not formulated in relation to evaluation objectives, and in relation to key evaluation criteria including relevance, and effectiveness. • Satisfactory – conclusions are consistent with data and well based on findings as well as meeting evaluation objectives and/or answering key questions. • Good – explanation of results consistent with data, conclusions well based on findings, conclusions represent actual insights into identification and /or solutions of important problems or issues. • Excellent –Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.
27. Quality of recommendations	<p>WHAT is the standard</p> <p>The preparation of recommendations needs to suit the evaluation process. Participation by stakeholders in the development of recommendations is strongly encouraged to increase ownership and utility. The planners and managers of the evaluation may decide to: include stakeholders in the creation of recommendations presented in the report; or may leave the consultation process for a separate stage after the report is completed.</p> <p>Recommendations should:</p> <ul style="list-style-type: none"> • follow logically from the evaluation findings and conclusions; • be relevant to the programme/project; • be clearly stated and not broad or vague; • be realistic and reflect an understanding of the commissioning organization and potential constraints to follow-up; • be prioritized with a timeframe for follow-up and suggest where responsibility for follow-up should lie; and • be formulated with participation of key stakeholders. <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Missing • Unsatisfactory – only partly meets first four areas above. • Satisfactory – adequately meets the first four areas above. • Good – Meets all of the areas above. This would usually be the result of some form of a formal management review and response to key evaluation findings and recommendations. • Excellent – Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

TOPICS	
TOR	Guide
<p>28. Quality of lessons learned</p>	<p>WHAT is the standard</p> <p>Lessons learned are contribution to general knowledge with implications for future action. They are learning from experience that is applicable to a generic situation rather than to a specific circumstance. The identification of lessons learned relies on three key factors: 1) the accumulation of past experiences and insights; 2) good data collection; and 3) a context analysis. Lessons learned should also try to involve stakeholders to some extent and in this sense not be entirely expert-led.</p> <p>Not all evaluations include lessons learned. If absent, this section should be skipped. A category 'missing' is therefore not provided.</p> <p>HOW is this standard applied</p> <ul style="list-style-type: none"> • Unsatisfactory -lessons learned are identified that are simply repetitions of basic knowledge about good programming, or which appear trivial. • Satisfactory -findings and conclusions of the evaluation are interpreted in relation to wider concerns, and lessons learned relate logically to the evaluations findings. • Good - findings and conclusions of the evaluation are interpreted in relation to wider concerns, lessons learned relate logically to the evaluation's findings, and there has been involvement of stakeholders in articulation of lessons learned. • Excellent - Evaluation meets all of the standards in a manner that can be considered best practice among the sample of evaluation reports.

Annex 3

Evaluations reviewed

AFRICA

Malawi - Final Evaluation for the Norad/SIDA/ UNFPA Supported Youth CBDA Initiative MLW/02/P01. *January 2004.*

Mozambique - Overview of Geração Biz Programme. Ministry of Education, Health, Youth and Sports in Partnership with NGOs and Youth Associations. *July 2004.*

Tanzania - Community-centered Lifeskills Education Programme to promote sexual and reproductive health among out of school youth. *January 2003.*

Swaziland - Report on Project Evaluation Contraceptive Logistics Management Information System Training Programme. *2002.*

Ghana - Report on the Evaluation of Project GHA/01/P08, Strengthening the Integration of Reproductive Health Services into Private Medical Practice. *January 2005.*

Ghana - Project Evaluation GHA/01/P06, Strengthening the Implementation of an Innovative and Culturally Sensitive Integrated Community Based Reproductive Health Service Delivery in the Upper East Region of Ghana. *20 December 2004.*

Burundi - Evaluation du projet BDI/03/P01, Renforcement des activités du Programme National de Santé de la Reproduction. *October 2004.*

Eritrea - Introduction and Use of the Female Condom by Selected Groups of Urban Women in Eritrea Project, An Evaluation Report, *December 2002.*

Eritrea - Final Evaluation on Emergency Reproductive Health : (ERI/00/P01 - Funded by US Bureau of Population, Refugees & Migration) & (ERI/01/P01 - Funded by the Dutch Government). *14 April - 7 May 2004.*

São Tomé - III Programme de assistance FNUAP au Gouvernement de São Tomé e Príncipe 1998-2001, Rapport de L'Evaluation Finale, *février - mars 2001.*

Ethiopia - Evaluation of the UNFPA PDS Sub-Programme of the Fifth Country Programme, Ethiopia 2003-2004, *October 8, - November 1, 2004.*

Botswana - Final Evaluation Report, Government of Botswana/UNFPA Third Country Programme (1998-2002), *May 2002.*

Burkina Faso - Revue a Mi-parcours du Programme de Cooperation 2001-2005. October 2003. Rapport de Synthese. Rapport du Sous-Programme Sante de la Reproduction. Rapport du Sous-Programme Population et Strategies de Developpement (PSD)

Zimbabwe - ZIM/00/P09 : Community and Workplace Based Peer Action Project for the Mutare Population. *April 2003.*

Zimbabwe - ZIM/00/P04 : Church and Community Based Reproductive Health and HIV/AIDS : An Independent Evaluation. *28 April 2003.*

Zimbabwe - ZIM/00P/02: Provision of HIV/AIDS/STI Voluntary Counselling and Testing, ZAPSO VCT project. *April 2003.*

AFRICA - REGIONAL EVALUATION REPORTS

Evaluation Report on the Activities of the UNFPA-funded SADC Census Support Projects (RAF/99/P03 and RAF/00/P03). *March 2003.*

Evaluation Mission on the Forum of African and Arab Parliamentarians for Population and Development. *May 2002.*

Strengthening Capacity to Operationalize Integrated Community-based Sexual and Reproductive Health Services in Sub-Sahara Africa (RAF/00/P05) - *November 2003.*

Assessment of SWAA at Regional and Country Level. *March 2003.*

Report, Evaluation of the Reproductive Health Project in the Refugee Camps of the Somali National Regional State (SNRS). *April 2004.*

Mali - Rapport de Mission - Republique du Mali. Evaluation Rapide du CERPOD. *May 2002.*

ASIA AND THE PACIFIC

DPRK - Report on Evaluation of the UNFPA Third Country Programme in Democratic People's Republic of Korea. 2003. *No date.*

China - Empowerment of Women through Improved Reproductive Health. CPR/98/P02. *November 2002 .*

India - Evaluation of Mobile Health Services in Earthquake affected Kutch District, Gujarat. *Sept 2003.*

India - Community Based Contraceptives Distribution Intervention in Rajasthan. *Oct. 2002.*

India - Evaluation Report on the Management of Integrated Population Development Projects in Maharashtra and Orissa by Dr. Nirmala Murthy and Dr. Alka Barua. *February 2002.*

Iran - Draft Report of Operational Evaluation 3rd UNPFA - Iran Country Programme, *May-August 2004.*

Philippines - Fifth Country Programme Mid-Term Review, Sub-programme Evaluation Reports: Reproductive Health and Reproductive Health SubProgramme - Adolescent Reproductive Health and Prevention and Management of Violence Against Women (VAW). *October 4, 2002*

Philippines - Fifth Country Programme Mid-Term Review. Advocacy Sub-programme.

Philippines - Fifth Country Programme Mid-Term Review. Population and Development Strategies Sub-programme.

Maldives - Evaluation Report of the UNFPA Second Country Programme for the Maldives (1998-2002). *January 25, 2003.*

Myanmar - Programme Assessment and Review, UNFPA Special Programme of Assistance to Myanmar (2002-2005). *25 October - 12 November 2004*

Indonesia - Report Documentation of Lesson Learned from UNFPA Project INS/01/P09: Establishment of District level Database. *October 2004.*

Indonesia - Evaluation Report of Population and Development Strategies Sub-Programme, As part of the Mid-term Evaluation of the UFPA Sixth Country Programme. *November 2003.*

Vietnam -UNFPA 6TH Country Programme Mid Term Review 2001-2005. July 2004.

Sri Lanka - Thematic Review of the Quality of Reproductive Health Services. Joint Review by Family Health Bureau, Ministry of Health and UNFPA, Sri Lanka. July 2004.

Sri Lanka - IEC and Advocacy in Support of RH - A Rapid Assessment. November 2004.

Sri Lanka - ARH Education in Schools - Mid Term Evaluation. 2004. **TOR**

ASIA AND THE PACIFIC - REGIONAL EVALUATION REPORTS

Evaluation Report of the Asia-Pacific Population Information Programme as funded by UNFPA during the Period 2000-2003. *June 2003.*

Evaluation of two regional projects (ARH and RH/IEC and Advocacy). *Oct. 2003.*

ARAB STATES AND EUROPE

Lebanon - Thematic Evaluation of the IEC and Quality of Care components under the Reproductive Health Sub Program (Projects LEB/97/P02 and LEB/98/P01). Partners for Development. *February 10, 2003.*

Bulgaria - Strengthening of the National Programme on Reproductive Health. BUL/00/P01. *November 2003.*

Belarus - Reproductive Health Information and Services Teenagers and Youth. BYE/00/P01. *21-28 March 2003.*

Belarus - Evaluation of project BYE/02/P01-Refinement of STIs/HIV/AIDS Prevention in the Belarus Armed Forces. *21-28 March 2003.*

Romania - Support For Reproductive and Sexuality Health Programme in Romania. ROM/00/P01. *May 2000-2004.*

Russia - Report on the Evaluation of the effectiveness of the project - Prevention of the spread of HIV among women involved in commercial sex in St. Petersburg. *November 2004.*

Algeria - Final Report on the external evaluation of project ALG/00/P01, Intégration du genre dans la santé reproductive : Prestation de services et plaidoyer; habilitation de la femme et éducation à l'égalité des droits. *December 2003.*

III. ARAB STATES AND EUROPE – REGIONAL EVALUATION REPORTS

Report on the Independent Evaluation of UNFPA Project **RMI98PO1** – Integration of Reproductive Health and Sexual Health into Boy Scouts and Girl Guides Programmes. *May 2002.*

LATIN AMERICA

Panama - Fortalecimiento de la Familia y Mejoramiento de la Salud Reproductiva del Pueblo Ngöbe. *November – December 2002.*

Nicaragua - Informe de Evaluación Final Externa NICO2P07 “Promoción de los Derechos, Equidad de Género y participación masculina para adolescentes y Jóvenes”. *November 2004.*

Nicaragua - Informe final de evaluación MSI “Acceso a IEC y servicios básicos de salud sexual y reproductiva para Adolescentes, Managua – Nicaragua”. *1 December 2004.* **LA4.Nicaragua**
- Informe final de evaluación MSI “Acceso a IEC y servicios básicos de salud sexual y reproductiva para Adolescentes, Managua – Nicaragua”. *1 December 2004.*

Ecuador - Evaluación del Proyecto: Salud Sexual y Reproductiva Para Adolescentes Frontera Sur – Ecuador. *December 2003.*

México - Final Evaluation Report of Mexico CP 1997-2001 conducted in 2002 (zip and maps files). *October 2002.*

LATIN AMERICA – REGIONAL EVALUATION REPORTS

Evaluation Report of Regional Project RLA/OOP08 and Interregional Project INT/OOP41 Agreed between UNFPA and JOICFP 2002-2003. *November 2003.*

Project RLA/00/P05: ICPD Following up and Monitoring through Civil Society Organizations. AETENEA NETWORK/Women’s Health Network for Latin America and the Caribbean (RSMLAC). No date.

INTERREGIONAL

Evaluation of the Averting Maternal Death and Disability Program, *July 26, 2004.*

The Evaluation of the Resource Flows Project, *May/June 2003.*

Evaluation Report on the Pilot Projects for Addressing Gender Based Violence in RH Clinics, *February 2004.*

Meeting the Development and Participation Rights of Adolescent Girls, Proj. INT01PAS. *31 March 2004.*

Annex 4

Interview guides

UNFPA Evaluation Quality Assessment Case study questionnaire

THE APPRECIATIVE INTERVIEW AND QUESTIONS

(May 10, 2005 final version)

1. Think of the time when you first heard or read the evaluation of this project or programme. What were your first impressions?
2. What made the evaluation so special?
3. What did you value most about the evaluation?
4. Thinking back at other evaluations you have read or come across, what insights can you share about these?
5. What do you think would make an effective evaluation?
6. If you could have three wishes for UNFPA for ensuring more of these successful evaluations, what would those wishes be?

METHODOLOGY FOR APPRECIATIVE INTERVIEWS

An interview with one other person:

1. Introduce yourself and say why you are here.
2. The interviewer should try to make the interviewee relax. You may want to simply describe what the meta evaluation is all about.
3. Briefly describe Appreciative Inquiry: Appreciative Inquiry seeks what works well and seeks to understand the elements of success so that they can be replicated.
4. Begin by asking the first question. Allow the person enough time to fully formulate and answer the questions before moving to the next question.
5. Jot down the answers using the first person I. Try to use the exact same words as the interviewee.
6. At the end of the interview, read back the story. Ask for any clarifications or additions.
7. Ask permission to use the interview in a case study. Persons can be anonymous.

An interview with three or more people

1. Introduce yourself and say why you are here.
2. The facilitator should try to create a relaxed atmosphere.
3. Explain Appreciative Inquiry: Appreciative Inquiry seeks what works well and seeks to understand the elements of success so that they can be replicated.
4. Ask the group to form groups of two or three people.
5. Assign roles: note taker, interviewer, interviewee. Tell the participants that they will have an opportunity to rotate roles after each of the interviews. Explain that the interview should be written in the first person, using the interviewee's same words.
6. The note taker should read back to the interviewee to make sure there are no omissions. Ask for any clarifications.
7. Rotate roles two more times until the group of three have each taken on a different role.
8. If time allows, read the stories in plenary.
9. Let the group identify convergences, common themes and gems (if time allows)
10. Ask permission to use the interviews in a case study. Persons can be anonymous.

A. EQA Case Study Interviews

1.0. UNFPA REP AND/OR DEPUTY REP:

Introduction

- How many years have you been with UNFPA?
- What is your current role and function within the office?

Appreciative Interview (Quality Evaluation Practices)

- 1.0. a). Think of the time when you first heard or read the evaluation of this project or programme? What were your first impressions? What made the evaluation so special? What did you value most about the evaluation? How is this evaluation been used? Is there anything you would do differently?
- 1.1. (Optional, if time permits) In looking back at your years of experience with UNFPA, can you remember an evaluation that stood out? What made it so special? What do you think constitutes good practice in evaluation?

Quality of Evaluation Reports

- 1.2. Is there an overall role for M&E within the office and/or with regards to projects?
- 1.3. What is the nature and type of resources allocated to M&E? (focal points for M&E, budget allocation)
- 1.4. (A3) Are UNFPA's current evaluation practices and approaches adequate and sufficient to help fulfill the requirements of UNFPA's new strategic directions? In light of the MYFF, are there currently themes or issues within UNFPA's mandate that are not being sufficiently evaluated?
- 1.5. What improvements are needed to support more consistent and better quality evaluations?

Quality of Evaluation Management

- 1.6. From your perspective, what role does or should evaluation play in the work of UNFPA?
- 1.7. How do you organize the M&E function within your office?
- 1.8. How much of an issue for you is evaluation quality? What in your view are currently the main quality gaps both in relation to the case study evaluation(s) and in other UNFPA evaluations?
- 1.9. Can you give any examples of where evaluation enabled you to improve programme quality and implementation (key purpose of evaluation)?
- 1.10. What is your approach to ensure that evaluation findings, conclusions and recommendation get their due management attention?
- 1.11. What kind of role do you see for M&E within SWAPs and Budget Support? Is attribution an issue? Are you considering more joint evaluations?
- (A3) What kind of role do you see for M&E in policy dialogue and advocacy? Are there any special M&E needs?

What kind of follow-up and response should be given to evaluations?

- 1.12. Thinking back on this and any other recent evaluations, how could evaluation have more influence for example on programming strategies, national population and development policies and strategies as well as UNFPA's organization-wide policies and approaches in key areas of our mandate?
- 1.15. What have been the effects or impact of the evaluation? How do you trace these?
- 1.16. (B3) Do you feel that UNFPA evaluations get used to influence policy dialogues?

Quality of available resources

- 1.17. (C2) Which improvements are needed to support more consistent and better quality evaluations? What are the main barriers to producing better and more consistent evaluations? Do you support or are you considering support to national evaluation capacity (either of counterparts or more generally)?
- 1.18. (C1) Are available human and financial resources sufficient for supporting evaluation work? What gaps exist? Any recommendations?
- 1.19. (C3) Do managers have the necessary tools and guidance materials to properly supervise and manage the evaluation process?
- 1.20. What kind of training in M&E should be made available? What are your suggestions for topics?
- 1.21. How strong is local/counterpart capacity in M&E? How could this be strengthened? Should UNFPA and UN in general be involved in strengthening local/counterpart M&E capacity?

Identification of quality evaluation practices

- 1.22. What instructive examples of quality evaluation results and practices can be highlighted to guide future evaluation efforts?
- 1.23. Do you have any examples of effective evaluation processes and management use? What factors seem to be associated with these success stories?

UNFPA Evaluation Manager

Introduction

- How many years have you been with UNFPA?
- What is your current role and function within the office?
- Is there an overall role for M&E within the office and/or with regards to projects?
- What is the nature and type of resources allocated to M&E? (focal points for M&E, budget allocation)

Appreciative Interview (Quality Evaluation Practices)

- 2.0. a) Think of the time when you first heard or read the evaluation of this project or programme? What were your first impressions? What made the evaluation so special? What did you value most about the evaluation? How is this evaluation been used? Is there anything you would do differently?
- 2.1. (Optional, if time permits) In looking back at your years of experience with UNFPA, can you remember an evaluation that stood out? What made it so special? What do you think constitutes good practice in evaluation?

Quality of Evaluation Reports

- 2.2. (A4) What does this evaluation mean for you in terms of what UNFPA needs to do to produce better and more consistent quality evaluations?
- 2.3. Can you give any examples of where evaluation enabled you to improve programme quality and implementation (key purpose of evaluation)?
- 2.4. From your perspective, what does UNFPA consider to be an adequate evaluation?
- 2.5. (A4) What are the ways UNFPA can improve evaluation quality?
- 2.6. What does UNFPA need to do to produce higher quality evaluations?
- 2.7. Have you participated in the dialogues on Results-based planning, monitoring and evaluation? What were the strengths and weaknesses of the dialogue? What did you learn from it?

Quality of Evaluation Management

- 2.8. What was your involvement with the evaluation? What was the evaluation process like - what were its strengths and weaknesses?
- 2.9. How was the evaluation ToR written up and finalized? Describe the negotiations with the consultant team concerning the TOR and budget? How did this impact on evaluation quality?
- 2.10. What were the key stakeholders of the evaluation? Were they involved in any step of the evaluation process? How was the decision on stakeholder involvement made?
- 2.11. Were the evaluation findings, conclusions and recommendations believable?
- 2.12. How did the evaluation successfully address the most relevant issues from your point of view? What insights did the evaluation provide?

What kind of Follow-up and Response should be given to Evaluations?

- 2.13. Was there a managed process for following up on the report and implementing its recommendations (e.g. recommendation tracking matrix, action plan)? How was the report disseminated (e.g. formal and informal briefings, key messages, workshops)?
- 2.14. (B5) How useful was the evaluation? Can you describe the many uses? Have any changes taken place as a result of the evaluation?
- 2.15. At what level has the impact been felt (e.g. Board, HQ dept, country office, field, government)? How was this tracked? What has the management response been (e.g. have any senior managers 'championed' the evaluation)?

Quality of Available Resources

- 2.16. (C3) How clear and relevant are the available guidance materials on evaluation? Do they support implementation of the new SD⁹? Are evaluation stakeholders and UNFPA staff able to make effective use of the guidance materials? Did you use the evaluation guidelines? How useful were these?
- 2.17. What kind of training in M&E should be made available? What are your suggestions for topics?
- 2.18. How do you locate good evaluators?
- 2.19. Was it possible to hire a consultant team with the required skill level and gender/national/international balance? What constraints did you face (e.g. short tendering period, lack of qualified consultants)? How did you overcome these?
- 2.20. Do staff and partners have a clear understanding of evaluation requirements?
- 2.21. (C1) Are available human and financial resources sufficient for supporting evaluation work? What gaps exist? Any recommendations?
- 2.22. (C2) What barriers and opportunities do staff and other relevant stakeholders perceive in the effort to produce better and more consistent evaluations?
- 2.23. How strong is local/counterpart capacity in M&E? How could this be strengthened? Should UNFPA and UN in general be involved in strengthening local/counterpart M&E capacity?

9. UNFPA's Strategic Direction (SD) approved in 2002 is an integrated organizational strategy, which identifies the priorities and strategic positioning of the Fund to contribute to the implementation of the ICPD PoA and the MDGs. According to the SD, UNFPA must demonstrate that it is an indispensable development partner in national and regional efforts to eradicate poverty and promote human rights. The following are three key areas of the Strategic Direction:

- Policy dialogue: UNFPA must be at the table when broader development policies and issues are discussed. The Fund must therefore build an evidence base, test innovative approaches and strategies and identify lessons learned to demonstrate that UNFPA funded interventions work and can be replicated on a broader scale for greater national impact.
- Advocacy: UNFPA must advance the ICPD agenda and clearly link it with the Millennium Development Goals (MDGs) and the PRS processes by addressing gender, reproductive health and rights, and population issues in the context of poverty reduction
- Partnerships: UNFPA should be part of innovative partnerships among development partners, governments and civil society organizations in which SWAps, sector reforms, PRS, the UNDAF and the CPAP are replacing individual projects as planning and programming tools

Source: MYFF 2004-2007 information Note Number 2, Strategic Planning Office

3.0 THE EVALUATORS

- 3.1. What was the evaluation process like – what were its strengths and weaknesses? What would you do differently next time? Was adequate time allocated for the evaluation?
- 3.2. Who were the key stakeholders of the evaluation? Were stakeholders involved in any step of the evaluation process? How was the decision about stakeholder involvement made?
- 3.2. Did you have a clear understanding of UNFPA evaluation requirements?
- 3.3. Were you provided with any evaluation guidelines? If so, were these useful?
- 3.4. What was your interaction with the UNFPA Evaluation Manager? How did you manage the evaluation manager?
- 3.5. Do you know of any follow-up to your report?
- 3.6. (A4) What improvements are needed to support more consistent and better quality evaluations?
- 3.7. Any suggestions for UNFPA that might help other evaluators in the future?

4.0. GOVERNMENT OFFICIALS AND COUNTERPARTS

- 4.1. What role did you play in the evaluation?
- 4.2. Did you participate in any of the evaluation steps?
- 4.3. What was the evaluation process like – what were its strengths and weaknesses? What would you do differently next time?
- 4.4. Were the evaluation findings, conclusions and recommendations believable?
- 4.5. Has the evaluation had an impact on the project or gone beyond the project? If so, how?.
- 4.6. Is there anything you think the evaluators could have done differently?
- 4.7. What kind of follow-up has there been to the evaluation? Are you satisfied with this?
- 4.8. (A4) What can UNFPA do to improve the quality of evaluations?
- 4.9. What is the current M&E capacity of your organization/department/ministry? How could this be strengthened?

5.0. EQA INFORMANTS

5.0. The Reference Group (HQ Managers and field-based)

- 5.1. In looking back at your years of experience with UNFPA, can you remember an evaluation that stood out? What made it so special?
- 5.2. What is your role in Evaluation?
- 5.3. What role do you feel evaluations should play in UNFPA?
- 5.4. Are there areas that need further strengthening?
- 5.5. What constitutes good practices in evaluation?

Quality of UNFPA Evaluation Reports?

- 5.6. (A4) What improvements are needed to support more consistent and better quality evaluations? Are there any barriers to produce better and more consistent evaluations?
- 5.7. (A3) Are UNFPA's current evaluation practices and approaches adequate and sufficient to help fulfill the requirements of UNFPA's new strategic directions? In light of the MYFF, are there currently any themes or issues within UNFPA's mandate that are not being sufficiently evaluated?

Quality of Evaluation Management and Attention to Follow-up?

- 5.8. (B1) Is UNFPA able to commission useful and timely evaluations, which address relevant issues and concerns for programme managers?

What Kind of Follow-up and Response should be given to Evaluations?

- 5.9. What kind of follow-up and response should be given to evaluations?
- 5.10. Do you feel that UNFPA evaluations get used to influence policy dialogues?

Quality of available resources

- 5.11. (C1) Are available human and financial resources sufficient for supporting evaluation work? What gaps exist? Any recommendations?
- 5.12. Do managers have the necessary tools and guidance materials to properly supervise and manage the evaluation process?

6.0. OTHER HQ INFORMANTS

6.0. Senior Directors/Programme Managers:

- 6.1. Think of an influential report you have read, you can remember or even you heard about? What were its qualities?
- 6.2. Are staff within the GDs monitoring evaluation results in any systematic way? Why not?
- 6.3. Any thoughts and suggestions you may have on UNFPA evaluation policy matters? How can we improve?
- 6.4. Are GD staff assigned any specific responsibilities re monitoring of evaluation results and follow through?
- 6.5. Do you have the time to read the evaluation reports?

How do you decide what to read and not to read? If you do read, what do you do about them?

7.0. TECHNICAL SERVICES DIVISION & CSTS:

- 7.1. From your perspective what role does or should evaluation play in the work of CSTs or technical advisors?
- 7.2. What are the main challenges and constraints you have been facing in improving evaluation quality at country level? How can these be overcome?
- 7.3. Is there any good evaluation practice that you can identify? What are the factors that lead to good evaluation practice?
- 7.4. What is your assessment of tools and resources being provided by HQ? Have you participated in dialogues, and if so, what is your assessment of these? What improvements are necessary to support better quality evaluations?
- 7.5. How do you use evaluations in your work? How is evaluation use and recommendation follow up monitored?
- 7.6. Should there be a stronger role for M&E? What additional resources are required?
- 7.7. Are evaluations useful for implementing strategic directions and to engage in policy dialogue?
- 7.8. What are some of the challenges in analyzing and applying lessons learned?
- 7.9. What work has the CST done in developing national capacities in RBM and planning, monitoring and evaluation? Do you see this as a future priority?
- 7.10. Do you have any recommendations for improvement of the quality of evaluation and its impact on programming?

Annex 5

Interviewees Listed Alphabetically

New York

PERSON INTERVIEWED	POSITION
Kate Alley	Evaluation Office, UNICEF
Fama Ba	Director, Africa Division, UNFPA
Olivier Brasseur	Director, DOS, UNFPA
Christine Bierring	Evaluation Advisor, UNFPA
Esteban Caballero	Programme Advisor to LAC, UNFPA
Linda Sherry-Cloonan	Deputy Director of Evaluation, UNFPA
Nural Alam	Evaluation office, UNDP
David Rider Smith	Evaluation Office, UNDP
Denise Deby	Senior Programme Officer, IDRC
France Donnay	TSD, UNFPA
Daniela Colombo	AIDOS
Patricia Guzman	Senior Technical Advisor, UNFPA
Sean Hand	Director of Personnel, UNFPA
Ayesha Iman	Culture and Ethics Cluster, UNFPA
Janet Jackson	DASE, UNFPA
Yaming Lin	Asia and the Pacific Division, UNFPA
Eduardo Mangas	LAC Division, UNFPA
Ian McFarlane	Special Assistant to the Deputy Executive Director, UNFPA
Pierre Moreau-Peron	Chief of Learning and Career Management Branch, UNFPA
Bill Musoke,	Deputy Director, Africa Division, UNFPA
Brendan O'Brien	Chief SPO, UNFPA
Julitta Onabanjo,	Technical Advisor, Young People, HIV/AIDS Branch, TSD, UNFPA
Marisela Padron	Director, LAC Division, UNFPA
Diego Palacios	Strategic Planning Office, UNFPA
Susan Pasquarella	SPO, UNFPA
Jiong Peng	APD, UNFPA
Prosper Poukouta	Africa Division, UNFPA
Bina Pradhan	Independent
Jean Quesnel	Head, Evaluation Office, UNICEF
Daniel Sala-Diakanda	Director, Africa Division, UNFPA
Elizabeth Santucci	Evaluation Office, UNICEF
Kunio Waki	Deputy Executive Director, Programme UNFPA

COUNTRY SUPPORT TEAMS

Ethiopia

PERSON INTERVIEWED	POSITION
Esther Muia	Officer-in-charge, CST Addis Ababa
Monique Rakotomalala	UNFPA Representative to Ethiopia
Friedl Van Den Bossche	HIV/AIDS Adviser

Mexico

PERSON INTERVIEWED	POSITION
Margaret Arilhs	CST Policy Advisor in Reproductive Health, Gender Rights
Beatrice Castellanos	CST, Sexual Reproductive Health Education Specialist
Sylvia Franco	CST, Sexual and Reproductive Health
Rogelio Fernandez-Castilla	Representative for Mexico, Director for Cuba and Dominican Republic, Director for CST
Jorge Sandoval	Assistant Representative

Thailand

PERSON INTERVIEWED	POSITION
G. Giridhar	CST Director and UNFPA Representative
Mere Kisekka	Adviser on Gender and Socio-Cultural Research
Chaiyos Kunanusont	Adviser on HIV/AIDS and STIs
Ghazy Mujahid	Adviser on Population Policies and Development
Josephine Sauvarin	Adviser on RH and FP Programmes
Pronchai Suchitt	Assistant Representative

COUNTRY OFFICES

Ghana

PERSON INTERVIEWED	POSITION
Dr. Edward Addai	Ministry of Health
Dr. Addico	NPO, RH
Doris Aglobites	Programme Officer, UNFPA program staff
Esther Apewokin	Ag. Executive Director, NPC
Jojo Baidu-Forson	Senior Research Fellow, UNU
Paul Dreguba	UNDP
Dr. Enyimayew	Consultant, RHI
Messeret Estefu	WHO Medical Officer
Dr. Melville George	UN Resident Coordinator
Bashirou Jahumpa	Assistant to Resident Coordinator
Robert Mensah	Coordination of African Youth Alliance, UNFPA program staff
Fatima Mrisho	Consultant in Reproductive Health
Prof. Nwuneli	Consultant, RHI
Dr. H. Odoi-Agyarko	Deputy Director, RCH Unit
Dr. Ofosu	Project Coordinator, SPMDP
Joana Opere	National Programme Manager/UN system gender programme
Susan Osam	NPPP
Mercy Osei-Konadu	NPO
Dr. Edith Tetteh	Evaluator
Emmanuel Tofatsi	ARR
Thomas A.N. Wobill	NPA

India

PERSON INTERVIEWED	POSITION
Hendrik van der Pol	UNFPA Representative
Venkatesh Srinivasan	Assistant UNFPA Representative
Sharareh Amirkhalili	Deputy UNFPA Representative
K.M. Sathyanarayana	UNFPA Technical Advisor

Iran

PERSON INTERVIEWED	POSITION
Mohamed Abdel-Ahad	UNFPA Representative
Soudabeh Amiri	Evaluator
Monire T. Basir	Programme Officer, UNFPA
Mohamad Mehdizadeh	Literacy Movement Organization, Ministry of Education

Lebanon

PERSON INTERVIEWED	POSITION
Dolly Bassil	Evaluator
Ms Joumana El Kadi	IEC project director
Mona Hammam	UNRC
Dr Mohamed Ali Kanaan	RH project Director
Amal Karaki,	Council for Development and Reconstruction
Asma Kurdahi	UNFPA Representative
Ramzi Namann	Evaluator
Arabia Osseiran	Evaluator
Ursula Rizk	Evaluator
Samira Suidan	Evaluator
Nadia Tewtel	SDSC (Planning)

Mozambique

PERSON INTERVIEWED	POSITION
Rui J.C. Albasini	Ministry of Youth and Sport
Martha Bazima	Programme Officer ASRH, UNFPA
Fernanda Bernardo Guianba	AMODEFA
Brigida Carvalho de Abreu	Maternal Mortality Reduction
Maria da luz Vaz	SRH Team Leader
Florbeca Fernandes	Assistant Representative, UNFPA
Leonardo Guirao	Pathfinder
Petra Lantz	Representative, UNFPA
Cristiano Matsinhe	Evaluator
Maria Molde	Programme Officer, ASRH, UNFPA
Alex Muiangar	Coordinator of Community Programme, AMODEFA
Julio Pacca	Country Representative, Chief Technical Advisor, Pathfinders.
Peter Reeh	Special Assistant, UNDP
Judy Senderowitz	International Evaluator
Jakob Sloth Madsen	JPO
Katia Taela	Evaluator
Antonio Tivane	Ministry of Education

Nicaragua

Person Interviewed	Position
Lotta María Aho	UNFPA Programme Staff
Camillo Antillon	Evaluator of NIC-02-PO7
Margaret Arilha	Advisor RH policies, Gender and Rights
Maria Bravo	Centro de salud
Jorge Campos	Population and Development, UNFPA
Beatrice Castellanos	Advisor Sexuality and RH Education
Dr. Luis Cuadra	Asistente Tecnico Principal
Maria de la Cruz Silva	Marie Stopes International
Sylvia Franco	Advisor Sexual and Reproductive Health of Youth
Hugo Gonzalez	UNFPA Programme Staff
Daniel Magnusson	UNFPA Programme Staff
Migdalia Molina	Marie Stopes International
Medea Morales	UNFPA Programme Staff
David Orozco	UNFPA Programme Staff
Chantal Pallais	NPPP, Adolescent Reproductive Health, UNFPA
Concepcion Puhiera	UNFPA Field Office Staff
Rosa Romero	AMUNIC, Executing Agency of NIC-02-PO7
Melvin Sotelo	Evaluator of NIC-01-PO3
Pedro Pablo Villanueva	Resident Representative UNFPA
Goya Wilson	Evaluator NIC-02-PO7
No Name Listed	- Coordinator of "Casa de Adolescentes" in Esteli - 12 youth, project NICO2P07. - 4 teenagers, Field Visit to Centro de Salud Project NICO2P03.

Sri Lanka

PERSON INTERVIEWED	POSITION
Kanthi Ariyaratne	Director, Health Education Bureau
Vinya Ariyaratne	Executive Director, Sarvodaya
Mr. Bandara	Health Education Bureau
Lubna Baqi	UNFPA Representative
Chithu Bandutilake	Family Health Bureau
Miguel Bermeo	UN Resident Coordinator
C. de Silva	Family Health Bureau
U. de Silva	Family Health Bureau
Chandani Galawaduge	National Project Officer, UNFPA
Anoma Jayatilaka	Family Health Bureau
Swarna Jayaweera	Evaluator
Vinitha Karunaratne	Director, Family Health Bureau
Thusitha Malalasekera	Health Education Bureau
N. Mapitigama	Family Health Bureau
Loshan Munasinghe	Family Health Bureau

PERSON INTERVIEWED	POSITION
Lakshman Senanayake	Evaluator
Janne Sykko	JPO
N. Vidyasagara	Evaluator
Mr. Wahid	Project Director, National Institute of Education
Malathi Weerasooria	Assistant Representative
Kusum Wickremasooriya	Evaluator

Vietnam

PERSON INTERVIEWED	POSITION
Pham Nguyen Bang	NPPP/PDS, UNFPA
Do Thi Minh Chau	PO/PDS, UNFPA
Duong Van Dat	PO/RH, UNFPA
Dinh Huy Duong	VCPFC
Nghiem Xuan Hanh	MOH
Nyugen Xuan Hong	NPPP/M&E, UNFPA
Ian Howie	UNFPA Representative
Hoang Thi Minh Huyen	MOFI
Le Thanh Huyen	NPPP/RH, UNFPA
Luu Quang Khanh	MPI
Phan Thi Le Mai	NPPP/RH, UNFPA
Quan Le Nga	Evaluator
Magali Romedenne	PO/Gender, UNFPA
Jordan Ryan	UN Resident Coordinator
Nguyen Huu Viet Tien	Evaluator
Tran Thi Van	UNFPA Assistant Representative

Annex 6

Background of team members

Tony Beck has a PhD in Geography from the University of London, and a BA and MA in English Literature from the University of Cambridge. He worked in India for four years carrying out research and as administrator for an India NGO, and since 1990 has been a consultant to 20 organisations working on over 50 contracts. Since 1994 he has focused on evaluation, specializing in evaluation design. He is the author of 2 books and 22 academic articles on development and evaluation issues.

Christina Bierring has a Masters in Socio-economic Development Planning from Roskilde University Center and a BA in Political Science from the University of Copenhagen, Denmark. She has 21 years experience in development planning, programme management, organizational development and evaluation. She has worked in the Latin America, Africa and Arab States regions including longer term field assignments to Peru, Zimbabwe and Benin, and as a Deputy Representative for UNFPA in India. As an Evaluation Advisor with UNFPA for the last 8 years, she has managed 2 large policy evaluations, conducted several evaluations and Country Office Management reviews world wide and authored the Programme Manager's Planning, Monitoring and Evaluation Toolkit.

Francoise Coupal has a Masters in Development Studies from Carleton University in Canada and a BA in Political Science from UCLA, United States. Ms. Coupal started her professional career working for the UNDP in Costa Rica and Niger and then the International Development Research Centre based in Ottawa. Ms. Coupal has lived and worked in Latin America, Africa and Asia. In 1992, Ms. Coupal founded her consulting firm Mosaic.net International focusing on participatory development, gender, special studies and evaluation. Ms. Coupal has undertaken over a dozen evaluations.

Scott Green has a BA in Political Science and an MA in International Development from Carleton University. He worked for three large international development agencies and in three countries in Africa for nine years for the United Nations World Food Programme (WFP). He has served as an Evaluation Officer at the WFP Headquarters in Rome, Italy from 2000 to 2004 and after that with the UNFPA Evaluation Office in New York. He has served as both team leader and team member on numerous evaluation studies.

Annex 7

Bias and limitations

1.1 THE EVALUATION TEAM

The evaluation team consisted of two independent evaluators and one UNFPA staff member from DOS as full team members, and one DOS staff member as a part time team member. There are various advantages and disadvantages of having a mixed internal-external team. The main advantages are that agency staff working on evaluations give profile to the evaluation, know their agency well, have a good sense of what recommendations are likely to be followed up, and can become an internal advocate for evaluation recommendations. Having agency staff as team members during Country Office visits is particularly useful as Country Office staff often feel more comfortable being interviewed by their own agency staff. The main disadvantage relates to the independence of the team. In this case one of the independent evaluators was the meta-evaluation team leader, which ensured independence. In terms of further potential bias, no meta-evaluator had been involved in any of the evaluations.

1.2 THE EQA SAMPLE

As noted in Annex 2, the sample of 60 evaluation reports is probably biased towards better quality evaluations. Divisions were asked to contribute evaluations that they thought were useful, and in any case it tends to be higher quality evaluations which are contributed to meta-evaluation exercises. If selection of evaluations had been non-purposive it is likely that overall rating would have been lower.

As in many meta-evaluation exercises, it was difficult to access an adequate number of evaluations. This was one of the reasons the sample was lowered from 80 to 60.

1.3 RATING AGAINST THE EQA MATRIX

Like any evaluation, meta-evaluations involve a certain degree of subjectivity in review of evaluations. However, meta-evaluators achieved a consistency in rating of 91 per cent, suggesting that the EQA matrix is a reliable tool (in the technical sense of a research or evaluation tool being likely to use the same process to achieve the same results repeatedly).