

D
O
S

Evaluation Quality Assessment: 2007-2008



Report No: UNFPA104

Issue Date: 12 February 2010

TABLE OF CONTENTS

TABLE OF CONTENTS	i
EXECUTIVE SUMMARY	1
INTRODUCTION	1
METHODS	1
Table 1: EQA Standards 2005, 2008, 2009.....	2
FINDINGS	4
Coverage	4
Table 2: Evaluations by UNFPA Region	5
Table 3: Submission rate by UNFPA Region	5
1.1) Objectives	6
1.2) Ethical Considerations	7
2.1) Terms of Reference	8
2.2) Assessment of Relevance	9
2.3) Assessment of Effectiveness	10
2.4) Assessment of Efficiency.....	11
2.5) Assessment of Sustainability	12
2.6) Assessment of Impact.....	13
3.1) Quality of Methods	14
4.1) Stakeholder Involvement.....	15
5.1) Gender Mainstreaming	16
6.1) Findings/Results	17
7.1) Conclusions	18
8.1) Recommendations	19
9.1) Executive Summary	19
9.2) Completeness of Reports.....	20
CONCLUSIONS.....	21
RECOMMENDATIONS	22
ANNEX 1	24
EQA STANDARDS	24

EXECUTIVE SUMMARY

UNFPA's Division for Oversight Services (DOS) has completed the 2009 Evaluation Quality Assessment (EQA) reported herein as part of its oversight and quality assurance responsibilities. Evaluations completed in 2007 and 2008 were reviewed and results compared to those from previous EQA exercises completed in 2005 and 2008.

Major findings include a marked decline in the number of evaluation reports submitted for the EQA in 2009 compared to the 2005 and 2008 EQA. Overall, quality remains generally poor with from 33-85% of evaluations missing or providing inadequate assessment of at least one of the 5 OECD/DAC evaluation criteria. Improvement was noted in the assessment of gender mainstreaming with over 50% of reviewed evaluations meeting expectations with regard to this standard for the first time since DOS began EQA exercises. Standards addressing reporting of ethical considerations in the design and conduct of the evaluation and stakeholder involvement have been added to the 2009 EQA.

Management is recommended to implement specific business process changes detailed in the report to improve both the coverage and quality of evaluation work done by UNFPA's country offices, regional offices, and headquarters units.

INTRODUCTION

This Evaluation Quality Assessment (EQA) has been completed by the Division of Oversight Services (DOS) as part of its oversight role. This assessment follows previous assessments in 2005 and 2008. Following Executive Board (EB) decision 2009/18, the EQA exercise functions as an explicit input into compliance with para 9 of the decision requesting that the DOS Director's biennial reports on evaluation address *inter alia* quality of evaluations.

METHODS

For comparability, the 2009 EQA built on the work from 2005 and 2008. The 2009 EQA was insourced and completed largely by DOS staff (previous EQA work had been largely outsourced to consultants). DOS staff used this opportunity to streamline the EQA process to be more relevant to two key target audiences: those who do evaluation and those who would use evaluation results to inform future planning and programming.

Fifteen standards were identified from previous work and similar assessment exercises by other UN organizations. The OECD/DAC evaluation criteria (relevance, effectiveness, efficiency, sustainability and impact) remain central to the 2009 EQA as they were in the 2005 and 2008 EQAs. Table 1 summarizes the standards used in the

2005, 2008, and 2009 EQAs. For the 2009 EQA, normative statements corresponding to ‘below expectations’, ‘meets expectations’, and ‘above expectations’ were defined for each standard using accepted norms from UNEG and others. These are summarized in Annex 1.

Table 1: EQA Standards 2005, 2008, 2009

DOMAIN	2005 STANDARDS	2008 STANDARDS	2009 STANDARDS
OECD/DAC evaluation criteria (relevance, effectiveness, efficiency, impact & sustainability)	Assessed for all evaluation reports (2005: Standards 21-25) (2008: Standards 6a, 6c, 6d, 6e & 6f) (Annex 1: Standards 2.2 through 2.6)		
Terms of Reference	Multiple aspects assessed (2005 & 2008: Standards 1a-1f)		Assessed holistically using a single standard (Annex 1: Standard 2.1)
Gender, Human Rights & Cultural Issues	Assessed via separate standards for gender mainstreaming, (2005: Standard 13, 2008: Standard 5a), human rights (2005: Standard 15, 2008: Standard 5b), and cultural issues (2005 only: Standard 17)) -cultural issues standard dropped after 2005; human rights standard defined in terms of identification of duty bearers and rights claims (2005, 2008) used to inform assessment of stakeholder involvement in 2009		Focus on gender mainstreaming (Annex 1: Standard 5.1)
Stakeholder Involvement	Assessed through multiple overlapping standards including human rights (see above), roles of stakeholders (2005 only: Standard 11), extent of beneficiary involvement (2005 only: Standard 12) capacity development (2005 only: Standard 14), national ownership (2005 only: Standard 18), partnership (2005 only: Standard 19)		Context-specific assessment of stakeholder involvement (Annex 1: Standard 4.1) with consideration of national ownership and partnership where relevant
Objectives & Ethical Considerations	Assessed via overlapping standards for purpose (2005 only: Standard 4), clear definition of evaluation users (2005 only: Standard 5), clarity of objectives (2005		Assessed via two standards: objectives (Annex 1: Standard 1.1) and

DOMAIN	2005 STANDARDS	2008 STANDARDS	2009 STANDARDS
	only: Standard 6) and consideration of ethics (2005 only: Standard 8); no evidence of assessment in 2008 EQA		ethical considerations (Annex 1: Standard 1.2)
Quality of Methods	Assessed via standards for quality of methods (2005: Standard 7, 2008: Standards 4a & 4b) and evaluator bias (2005: Standard 9) or bias & limitations (2008: Standard 4c)		Assessed via a single standard (Annex 1: Standard 3.1) including triangulation and attention to bias as integral to quality of methods
Findings	Not assessed		Assessed via single standard focused on linkage of findings to methods used (Annex 1: Standard 6.1)
Conclusions	Assessed via single standard (2005: Standard 26, 2008: Standard 7, Annex 1: Standard 7.1)		
Recommendations	Assessed via single standard (2005: Standard 27, 2008: Standard 8, Annex 1: Standard 8.1)		
Executive Summary	Assessed via single standard (2005: Standard 3)	Assessed by Y/N checklist item (2008: Standard 2a)	Assessed via single standard as in 2005 and included in completeness checklist (Annex 1: Standards 9.1 and 9.2)

For comparability with the five point scale used in previous EQAs, scores of 'satisfactory' and 'good' from 2005 and 2008 were combined and deemed equivalent to 'meets expectations' in the 2009 assessment. The reviewers concurred that establishing a consistent basis for distinguishing between 'satisfactory' and 'good' would introduce potential bias and complexity out of proportion to the potential benefits for the EQA target audiences. In addition, the 2008 EQA report did not include an overall assessment of the reviewed evaluations, providing separate results by year of evaluation for 18 evaluations from 2006 and 16 from 2007. The scores for the 2008 EQA presented in this report are thus a weighted average (by numbers of evaluations in each year) of the 2006 and 2007 results from the 2008 report. This weighted average provides 2008 results calculated in a way that enables comparison to the 2005 and 2009 results.

As UNFPA maintains no central evaluation registry or other recording system, evaluations were identified through a mid-2009 memo sent to all country office (CO)

representatives, regional office (RO) directors, and headquarters (HQ) directors. The memo requested that any evaluation, or work that might be deemed evaluation, completed in 2007 or 2008 be provided to DOS. Two follow-up reminders were sent.

One hundred and twenty seven documents were submitted by COs. All submissions received were logged and categorized by country and region. Seventy submissions with no evidence of evaluation (e.g. baseline surveys, research papers) were excluded. From the 57 evaluations, nine midterm reviews were excluded and a sample of 37 evaluations was selected for review from the remaining 48 evaluations.

The sample was designed to represent submissions from four regions from which submissions were received, (two evaluations were received from a single CO in the fifth region after the final deadline), to limit to two the maximum number of evaluations per country in the sample, and to include evaluations in English, French, and Spanish. The sample was selected based solely on the CO submitting the evaluation and its language and without any knowledge of the content or quality of the evaluation. After the selection, four evaluations were noted to have been included in the sample used for the unpublished 2008 report. Review of the results of the assessment determined that this duplication was immaterial to the results for 2009 and these four evaluations were thus kept in the sample.

Each of the 37 evaluations was reviewed by two reviewers. For English and French language evaluations, each was reviewed by two of four DOS staff members. After each staff member completed his/her review, the two reviewers met and established consensus scores for each EQA standard. Spanish-language evaluations were all reviewed by the Regional Monitoring & Evaluation Adviser in UNFPA's Latin America and Caribbean Regional Office and a locally-retained Spanish speaking consultant using the same standards and consensus approach to determine final scores.

FINDINGS

Findings for coverage and then each of the 15 standards are presented in the sections that follow. For each of the standards, the domain is briefly described ('Description'). Details of the normative statements for 'below expectations', 'meets expectations' and 'above expectations' for each standard are provided in Annex 1. The last finding reports on the completeness of evaluation reports.

Coverage

While the request for evaluations was sent to all CO, RO and HQ, the pattern of responses indicates that evaluations are submitted to the EQA process from a limited number of business units. No submissions were received from RO or HQ divisions. Among COs, table 2 below summarizes the numbers by region.

Table 2: Evaluations by UNFPA Region

UNFPA REGION	Number of CO submitting to EQA & Total Submissions	COs submitting at least one evaluation
Africa	12 countries 50 submissions 18 evaluations	Angola, Benin, DRC, Eritrea, Ethiopia, Gambia, Lesotho, Mauritania, Namibia, Senegal, Tanzania, Uganda
Arab States	2 countries 13 submissions 5 evaluations	Morocco, Sudan
Asia-Pacific	9 countries 41 submissions 13 evaluations	Bangladesh, Cambodia, Indonesia, Iran, Lao PDR, Pacific Islands, Sri Lanka, Timor Leste, Vietnam*
Eastern Europe & Central Asia	1 country 2 submissions 2 evaluations**	Russia
Latin America & Caribbean	7 countries 21 submissions 10 evaluations	Bolivia, Colombia, Ecuador, Haiti, Honduras, Mexico, Nicaragua

*: Vietnam CO submitted 10 documents, however none were deemed evaluation.

** : The submission from the Russian Federation CO was received too late to be included in the sample.

Submission rate by region was less than 50% in all regions, ranging from 5-39%. Results are summarized in table 3.

Table 3: Submission rate by UNFPA Region

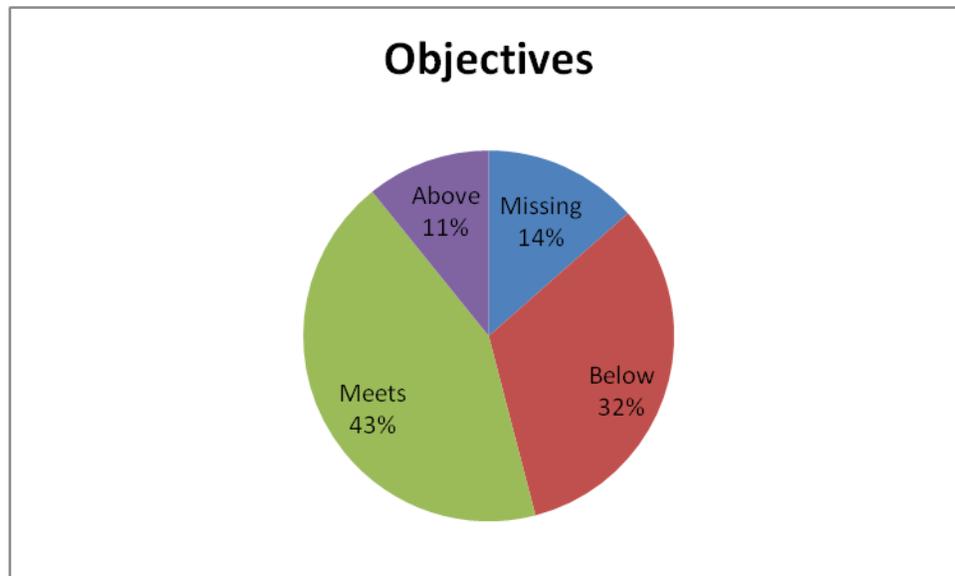
REGION	NUMBER OF UNFPA COUNTRY OFFICES	Proportion of COs submitting material to EQA
Africa	45	27%
Arab States	14	14%
Asia-Pacific	23	39%
Eastern Europe & Central Asia	20	5%
Latin America & Caribbean	21	33%

The sections that follow graphically summarize results for each standard and, where available, comparisons with 2005 and 2008. Comments following each figure are based on the reviewers' comments aggregated over the sample of evaluations. Guidance for evaluators is intended to provide practical direction to those who commission, manage or implement evaluations for UNFPA.

1.1) Objectives

Description: Evaluation objectives provide a clear statement of what the evaluation seeks to accomplish. Objectives can be detailed further in specific evaluation questions.

N.B.: This item was not comparably assessed in the 2005 and 2008 EQAs so results below are from the 2009 EQA only



In summary, 54% of evaluations were rated as meeting or exceeding expectations. A clear statement of objectives is essential to guide the evaluation work and also, for readers of the evaluation report, to enable them to identify whether the evaluation is likely to be relevant to their particular context.

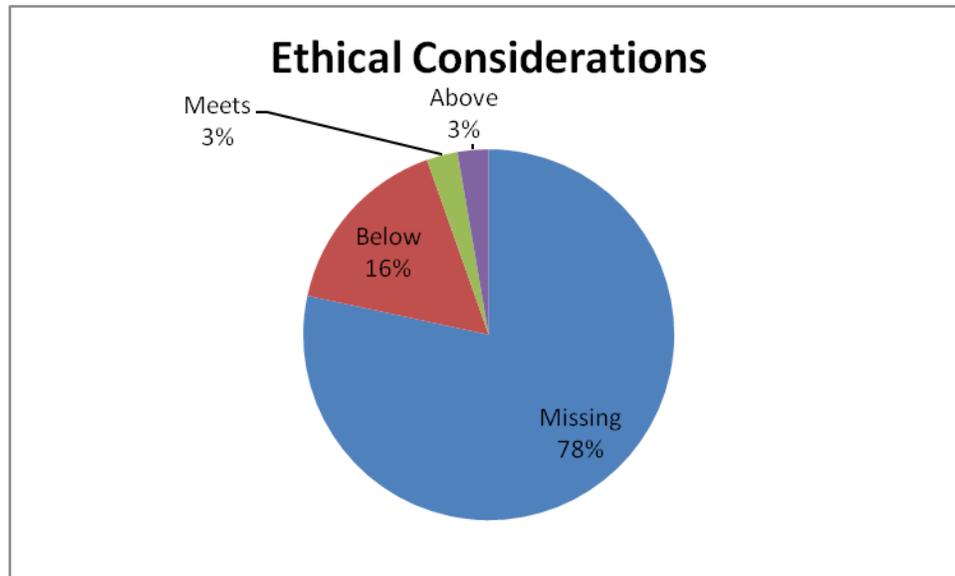
Guidance for Evaluators

- Optimal objectives are specific rather than generic
- Objectives should go beyond merely verifying that project or programme activities were completed.
- Objectives of evaluation (or 'evaluation questions') can be particularly effective if grounded in the objectives of the intervention or programme as determined during design of the intervention or programme.

1.2) Ethical Considerations

Description: The evaluation report describes mechanisms and measures that were implemented to ensure that the evaluation process conformed with relevant ethical standards including but not limited to informed consent of participants, privacy and confidentiality considerations

N.B.: This item was not comparably assessed in the 2005 and 2008 EQAs so results below are from the 2009 EQA only.



Ethical considerations include important aspects such as informed consent from participants and management of personally identifiable information. Raters differed in their views as to whether 'missing' constituted 'below expectations', but the striking finding from this analysis is that 94% of evaluations failed to meet expectations with regard to documenting such basic items as how informed consent was ensured.

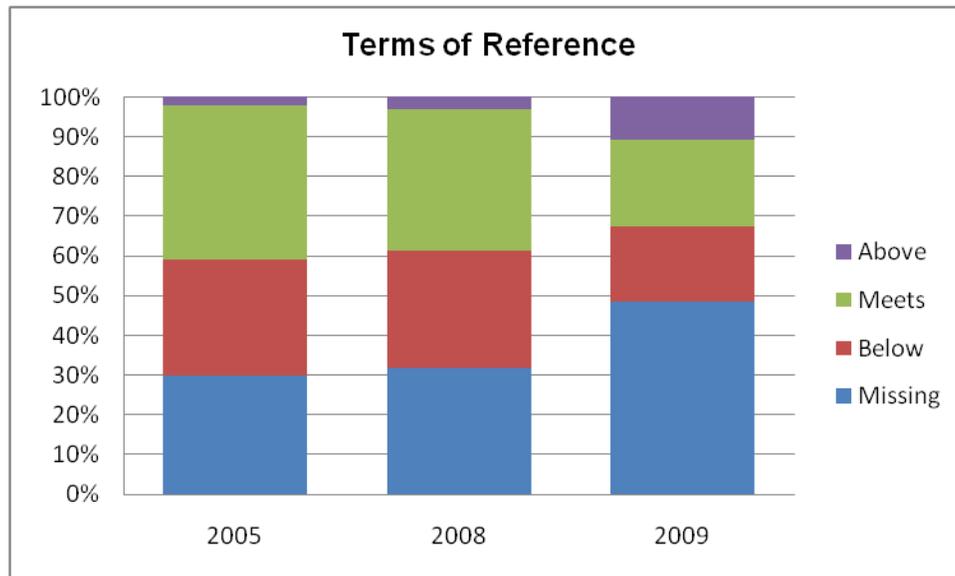
Guidance for Evaluators

- Minimum expectations for ethical considerations would include documentation of consent procedures where beneficiaries or members of the public are surveyed.
- Brief descriptions of confidentiality provisions should be provided where personal information or personal health information is used in the evaluation or the evaluation report
- Institutional review board or research ethics approval as appropriate to the nature of the evaluation and context should also be mentioned. If no such approval was sought or deemed relevant, stating that reduces uncertainty for the reader.

2.1) Terms of Reference

Description: The terms of reference (ToR) clearly describe the evaluation's intended scope and focus –either by defining the main evaluation questions, and/or listing main objectives.

•



For Terms of Reference (ToR), data from previous EQA efforts enable comparison of trends over time. Half of the 2009 sample did not include any ToR and this is a decline from previous performance. Among reports including ToR, no significant improvement was noted over previous performance.

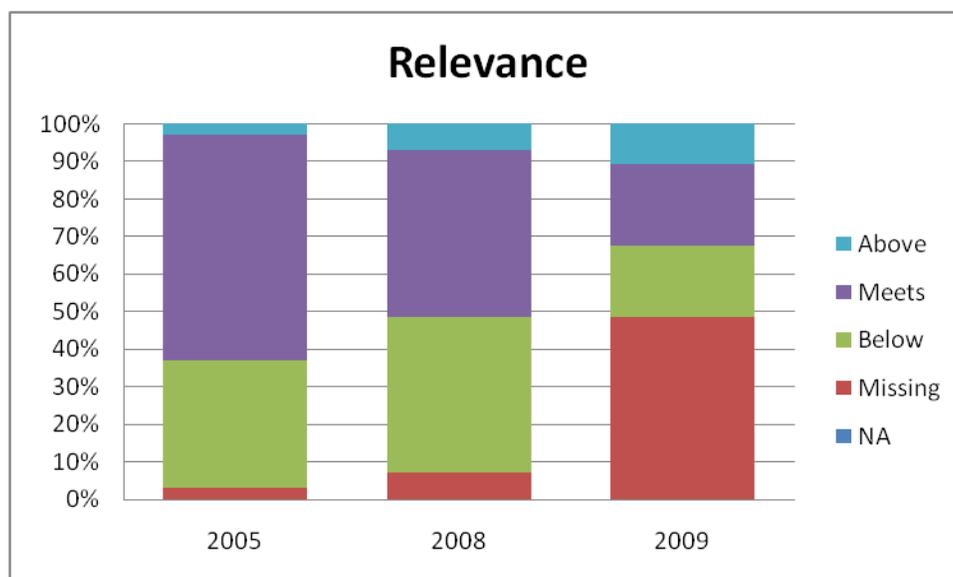
Guidance for Evaluators

- Quality ToR provide specific detail regarding the evaluation questions and the methods proposed to generate data to address each evaluation question.
- Efforts should be made to consider and reference all 5 OECD/DAC evaluation criteria: relevance, effectiveness, efficiency, sustainability, and impact, and how these will be assessed in the evaluation.
- Optimal ToR will also include a project plan listing dates and deliverables and be included as an Annex to the evaluation report

2.2) Assessment of Relevance

Description: Assessment of programme/project relevance examines the degree to which the outputs/outcomes are in line with national needs/priorities, UNFPA priorities, and relevant to stakeholders.

N.B. Many of the evaluations in the 2009 EQA were evaluations of UNFPA country programmes. In this case, relevance was often not directly assessed and may have been assumed to be implicitly addressed in the structure of the programme.



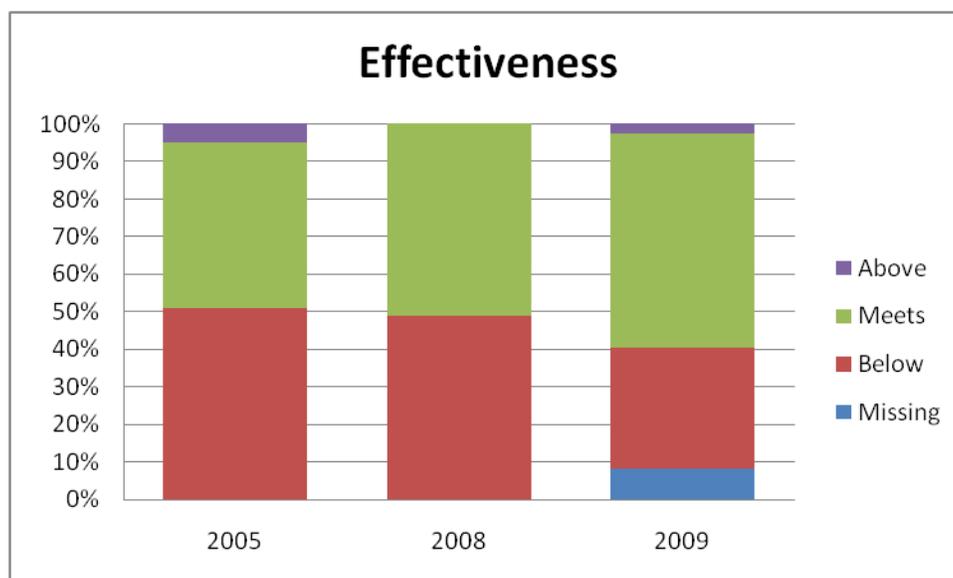
With regard to relevance, reviewers in 2009 appear to have used a different interpretation than in previous years due to the assumption of implicit relevance assessment made for many UNFPA country programme evaluations in the past. Regardless of interpretive differences, the combination of 'missing' and 'below expectations' has steadily increased since 2005.

Guidance for Evaluators

- Assessing relevance is generally a matter of assessing the alignment of the project or programme under evaluation with some combination of the country's national development objectives, UNFPA's strategic plan and ICPD commitments.
- Approaches to assessing relevance may include document review, interviews with other development partners and consideration of alternative interventions.

2.3) Assessment of Effectiveness

Description: Assessment of effectiveness examines the extent to which a programme/project achieves its planned results (outputs and outcomes)



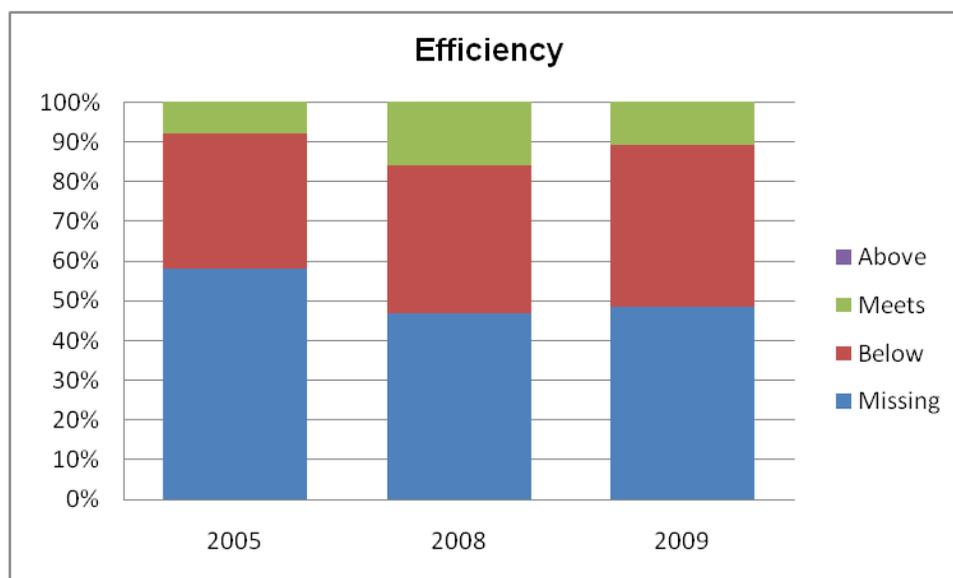
Assessment of effectiveness ideally includes consideration of how outputs/results or outcomes will be verified. In many cases, however, assessment of effectiveness amounts to little more than verification of activities. Performance in 2009 is essentially unchanged from previous years. 40% of the evaluations' included in the 2009 EQA sample did not include any assessment of effectiveness that would meet minimum expectations.

Guidance for Evaluators

- Evaluations can begin assessment of effectiveness with verification of whether planned outputs and ideally, outcomes were achieved. Assessing how the programme or project being evaluated contributed to or was responsible for the changes in outputs or outcome indicators requires both output and outcome indicators. Verifying only that activities were completed is rarely if ever sufficient for an evaluation.
- Outcome and output assessment is optimally based on data drawn from trusted, impartial or objective sources but will often include subjective elements such as partners' impressions or opinions.

2.4) Assessment of Efficiency

Description: Evaluation assesses efficiency, linking outputs to expenditures/resources and assessing whether this occurred as economically as possible.



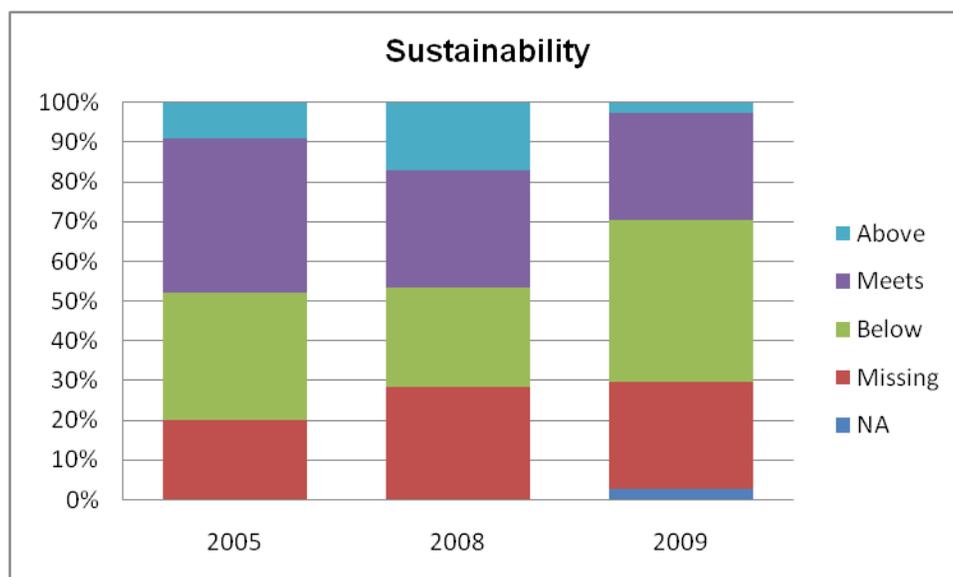
Efficiency assessment is critical for scaling up promising interventions or programmes. Evaluating efficiency can help identify whether scaling up requires only adding resources in a predictable fashion or whether the intervention or programme itself may need to be redesigned to be affordable at scale. While the proportion of evaluations lacking any assessment of efficiency declined slightly over the period 2002-2009, no concomitant improvement in the proportion meeting or exceeding expectations occurred.

Guidelines for Evaluators

- Assessing efficiency requires some data on resources used and the costs and quantities of those resources. Where UNFPA business units (CO, RO, HQ) use contractors to complete evaluations, contracted evaluators should be provided access to UNFPA's data on resources used and costs. This should be included in ToR for most if not all evaluations.

2.5) Assessment of Sustainability

Description: Sustainability assesses the extent to which programme/project results are likely to continue/remain after termination of external assistance.



Performance with regard to assessing sustainability is largely unchanged from previous EQA exercises. Less than one-third of evaluations meet or exceed expectations with regard to assessing sustainability, without which the development aspects of UNFPA's work are likely imperiled.

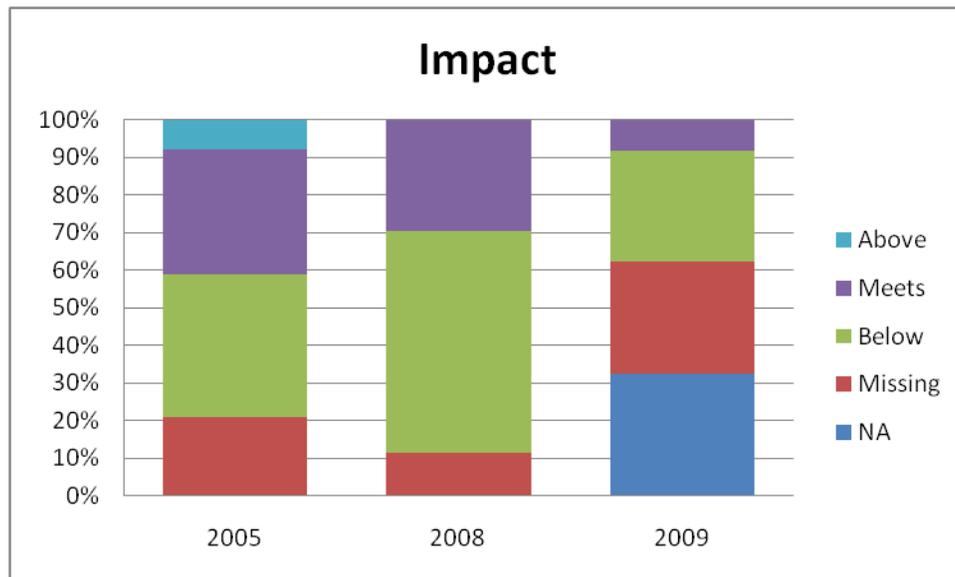
Guidance for Evaluators

- Sustainability is often overlooked in a programme context because programmes of cooperation are expected to continue. Nevertheless, evaluators can examine the evolution of roles of both UNFPA and implementing partners over the time period of the evaluation for evidence of durable capacity development and increasing partner ownership and direction of activities

2.6) Assessment of Impact

Description: Impact assesses positive and negative long-term effects which may be economic, socio-cultural, institutional, environmental, technological or other effects. Considerations of scale are also important when evaluating pilot interventions or projects

N.B. Reviewers for the 2009 EQA concurred in the view that impact assessment was rarely if ever possible, since evaluations, particularly of UNFPA country programmes occur very shortly after or even before all programme activities are completed. Thus, the 2009 EQA results report a substantial increase in the 'NA: not applicable' category.



Comparability of the 2009 results to 2005 and 2008 results appears to be limited by differing interpretations of the standard. The most likely explanation is that in previous EQA exercises, reviewers considered a more flexible notion of impact along the lines of 'change in outcome'. For 2009, the EQA reviewers applied the OECD/DAC definition of 'The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators. The examination should be concerned with both intended and unintended results'.¹ Applying the 2009 interpretation, only 10% of evaluations make an acceptable effort to assess impact.

Guidance for Evaluators

- Evaluators are encouraged not to overlook impact evaluation or dismiss it as 'long-term' impact. Examining programme or project theory and/or logic models to identify how the activities under evaluation foresaw impact can guide evaluation of impact, even if the 'long-term' effects remain in the future.

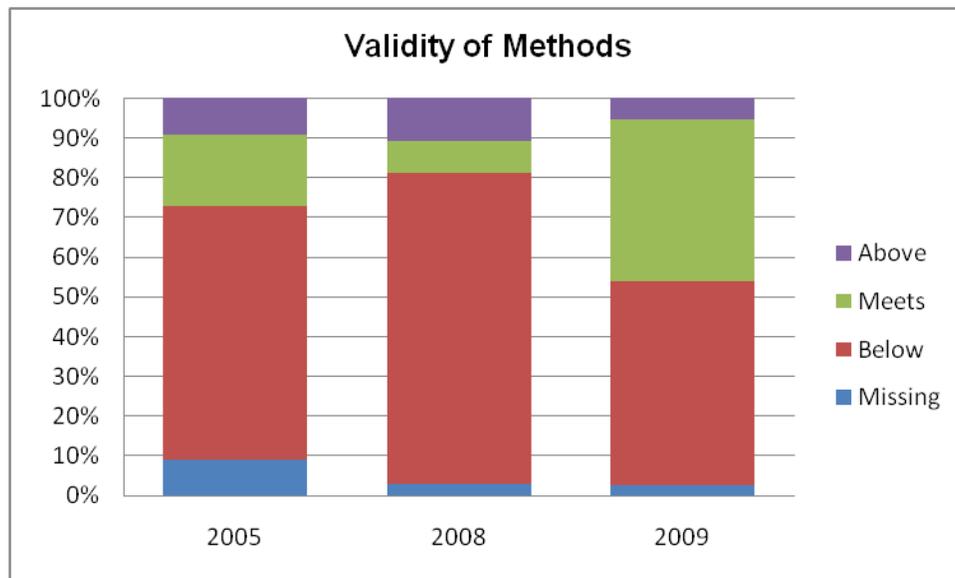
¹ http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html

- Considering scale is important for any pilot intervention or project since an intervention that improves outcomes but is simply too complex or too expensive may well need redesign lest scaling-up be impossible due to resource constraints. Scale often can be assessed by determining how many people were reached or impacted out of how many i) were intended to be reached and ii) would have benefitted.

3.1) Quality of Methods

Description: Methods are valid i.e. focused on and logically linked to evaluation objectives and questions, consistent with good practice and include, where appropriate explicit efforts to test counterfactuals and triangulate among data sources, and efforts to control bias and/or acknowledge limitations due to uncontrolled bias are implemented and described.

N.B.: In previous years, these three elements were treated separately. In light of the emphasis in the 2009 EQA on providing information useful to readers, they have been combined with a primary emphasis on the validity of the methods, since valid methods can be understood as those making provision for triangulation and bias control. For comparison purposes, results reported for ‘good practice’ have been used for the 2005 & 2008 EQA.



An increasing proportion of evaluations reviewed in the EQA are including descriptions of the methods used. However, over 50% of reviewed evaluations had methods sufficiently weak as to call into question the validity of any findings and thus, conclusions and recommendations. Selecting and using valid methods are foundational to quality evaluations since weak methods or methods proposed but not implemented provide a weak basis for findings, conclusions and recommendations.

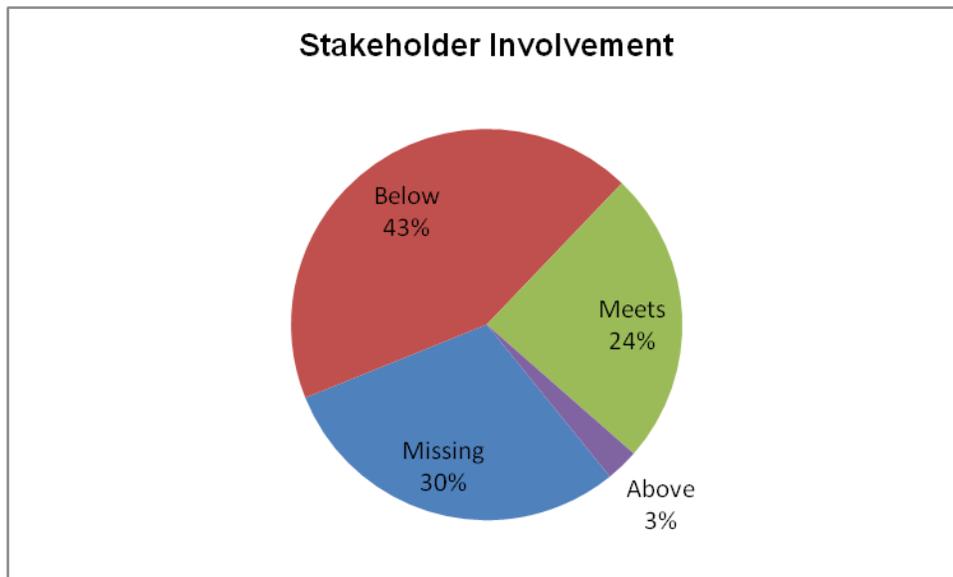
Guidance for Evaluators

- At a minimum, an evaluation report needs to describe the methods used and provide evidence that the methods proposed were implemented and generated evidence that supports the findings. Proposed methods should be linked to evaluation questions (i.e. this question will be addressed using this method).
- Particular attention is needed to bias control and triangulation. Triangulation contributes to bias control since drawing on multiple sources of information and actively considering and seeking to validate alternative explanations for results can help reveal and thus, control bias.
- Where methods proposed in the ToR are subsequently not used in the actual evaluation, evaluators should explain this in the evaluation report text (e.g. not practical, too expensive).

4.1) Stakeholder Involvement

Description: Consideration is given to stakeholder involvement and report is clear about rationale for and level of stakeholder involvement

N.B.: This item was not comparably assessed in the 2005 & 2008 EQAs so results below are from the 2009 EQA only.



Stakeholder involvement should be tailored to the context of each evaluation and may include participating in design (questions/objectives, methods, data collection instruments), collecting data, analyzing data, or developing recommendations. Survey participation (i.e. as a respondent alone) would generally fall short of the stakeholder involvement necessary to support national ownership of development cooperation and national evaluation capacity development. The 2009 results demonstrate substantial

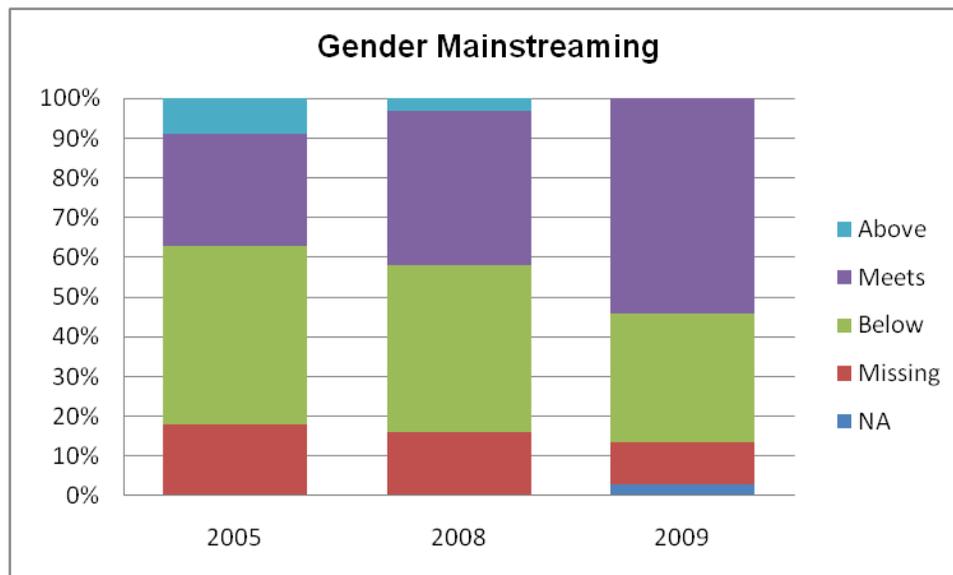
scope both for deepening stakeholder involvement and for documenting that involvement in evaluation reports.

Guidance for Evaluators

- The broader development context increasingly emphasizes national ownership. For evaluations, this requires meaningful involvement of national counterparts in the conception, design, execution and follow-up of evaluations. Considering how to involve stakeholders at each of these four steps is a minimum expectation.
- Participation in surveys and interviews may be critical to gather relevant information for an evaluation but is not the same as 'involvement'. Where surveys or interviews are used, results should be presented on non-responders as well as survey participants.
- Follow-up to evaluation findings and recommendations with stakeholders completes the involvement cycle and builds shared ownership.

5.1) Gender Mainstreaming

Description: Where relevant, evaluation assesses extent to which data disaggregated by gender were used for planning and assessing programme/project and extent to which programme/project promoted gender mainstreaming.



The 2009 EQA sample was marked by the lowest level of 'missing' for this element since the EQA exercises began. In addition, more than 50% of evaluations were deemed to have met or exceeded expectations with regard to gender mainstreaming.

Guidance for Evaluators

- Gender mainstreaming in different development contexts can range from using gender-disaggregated data to comprehensive gender-informed interventions. For

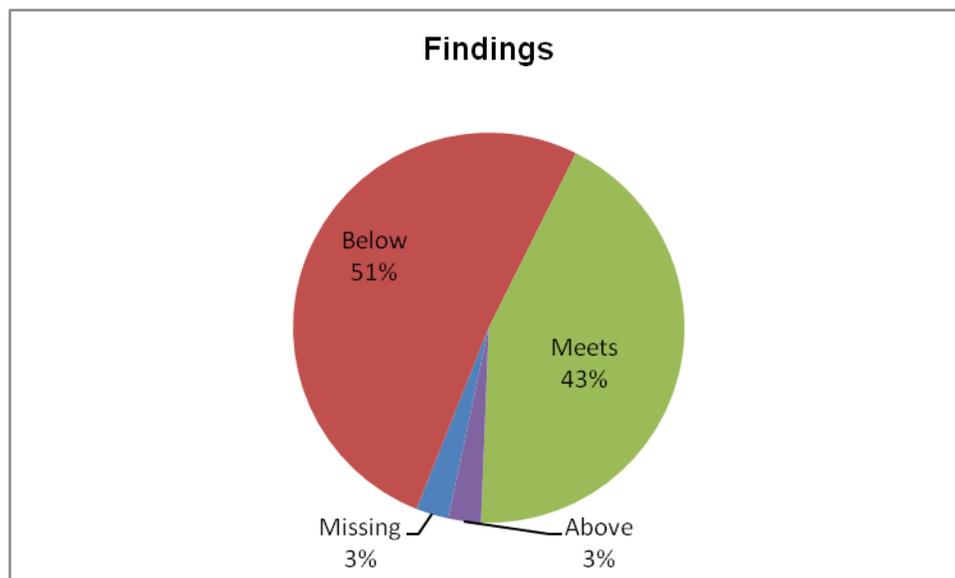
each evaluation question, evaluators are encouraged to ask what the relevant gender mainstreaming considerations are and how each will be assessed.

- Where outcomes are assessed by evaluators, designing data collection instruments to permit disaggregation of results by gender must be done before data collection commences.

6.1) Findings/Results

Description: Findings/Results are analyzed in terms of outcomes or impacts, including cost analyses, and reasons for accomplishments/difficulties are identified and supported by analysis.

N.B.: This item was not comparably assessed in the 2005 and 2008 EQAs so results below are from the 2009 EQA only.



Findings are foundational for both valid conclusions and actionable, relevant recommendations. Relevant findings must also be grounded in the methods used and data gathered by the evaluators. In the 2009 EQA, over half of evaluations failed to meet these basic expectations. The most common deficiency was the statement of findings or results unsupported by the data in the evaluation report. While expert opinion can provide a reasonable basis for findings and recommendations, the evaluation report needs to provide assurance to the reader that this opinion is at least informed by the experience of the intervention or programme being evaluated.

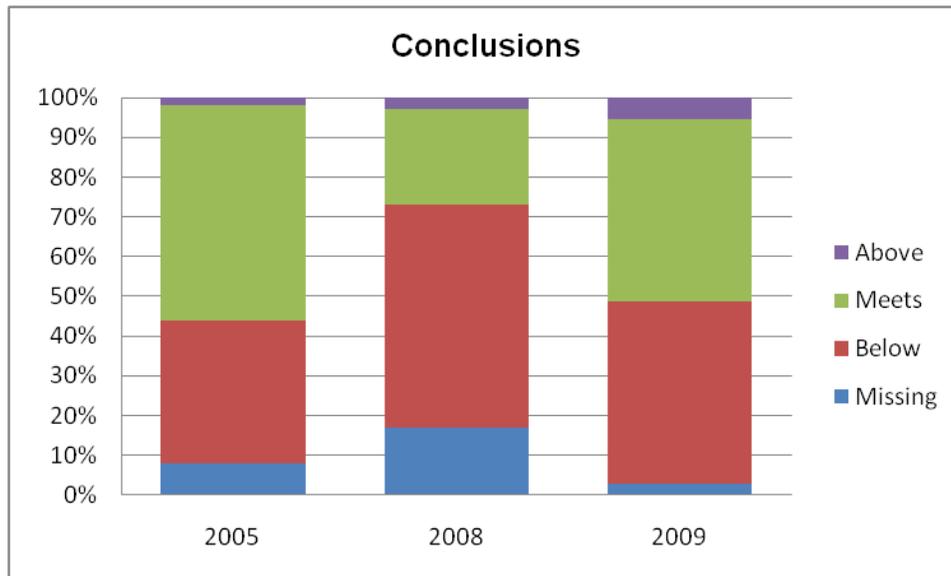
Guidance for Evaluators

- Findings focused primarily on activities or outputs need to be supplemented by findings on outcomes. If outcomes were not assessed during the evaluation, explanation for this omission can be useful.

- Each finding should build on evidence or data presented in the evaluation report. If a finding seems important but lacks supporting evidence, asking on what grounds this finding is deemed to be important can identify potential sources of information or evidence that may merit inclusion in the evaluation report.

7.1) Conclusions

Description: Conclusions flow logically from and are supported by evaluation findings and address issues of significance to the programme/project as scoped by the evaluation questions/objectives.



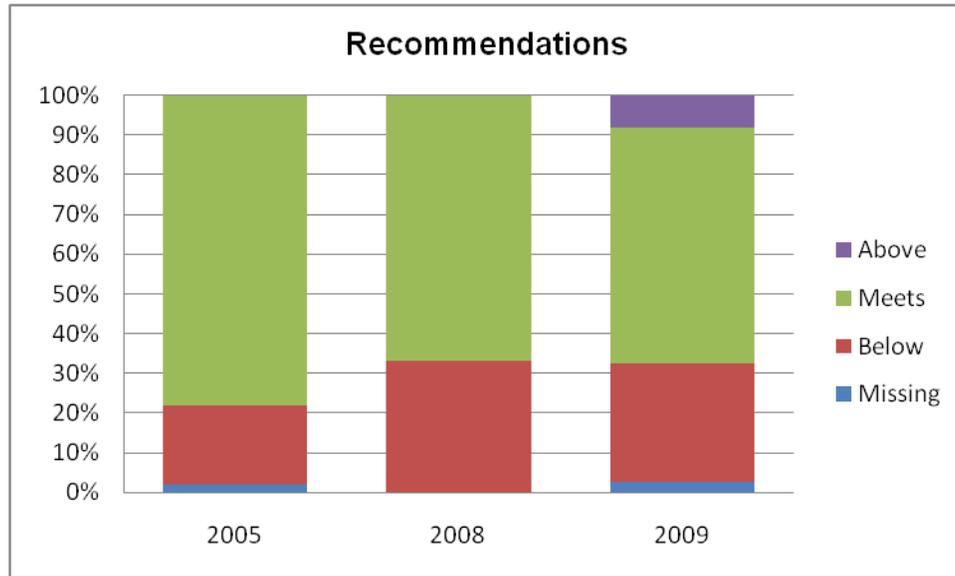
In 2009, more than 50% of the EQA sample met or exceeded expectations on this item. However, this improvement over 2008 simply returns to the level of performance reported for the 2005 EQA.

Guidance for Evaluators

- Evaluators and evaluation managers are encouraged to distinguish between conclusions (interpretations of findings), recommendations (actions proposed as a consequence of findings and conclusions) and lessons learned (knowledge or processes potentially applicable to other settings or programmes).
- Conclusions are the logical linkage between the findings or results of the work done by the evaluators and the recommendations made by the evaluators.

8.1) Recommendations

Description: Recommendations are supported by data analyses, findings and conclusions, and are clearly stated with specific details of who is recommended to do what by when.



EQA assessments of recommendations report that most evaluations have generally met or exceeded expectations and this trend continues in the 2009 sample. The proportion of reports missing recommendations or failing to meet expectations remains at approximately one-third.

Guidance for Evaluators

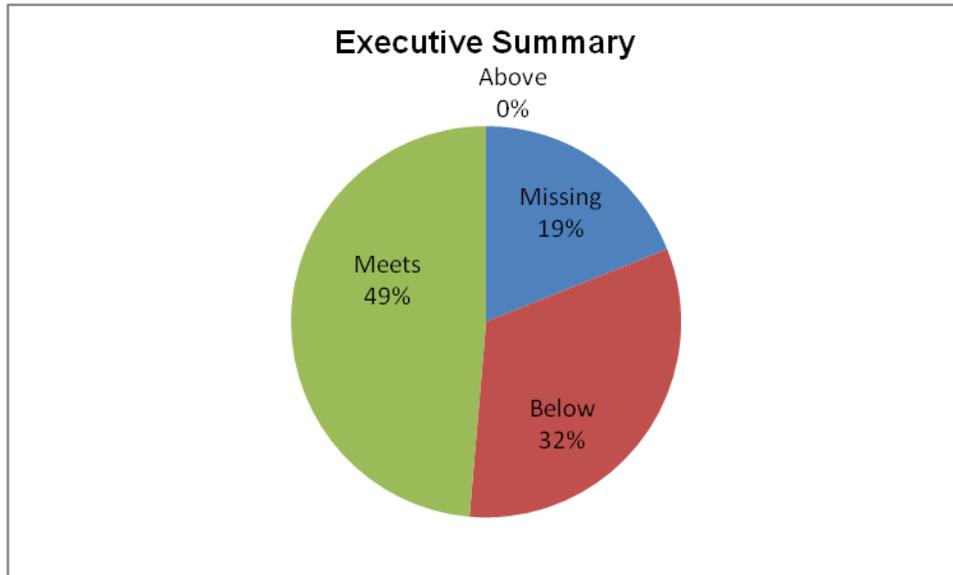
- Effective recommendations need to include specific actions to be undertaken by specific parties and completed by specific deadlines.
- Prioritizing recommendations can increase their effect.
- Where surveys or interviews are used, asking survey or interview participants for recommendations and presenting these unaltered as 'participant recommendations' to complement the evaluators' recommendations can be valuable.

9.1) Executive Summary

Description: Executive Summary is concise, readable and can stand alone without reference to the rest of the report.

N.B. In 2005 & 2008, the assessment of the executive summary was done in different ways (see Table 1). In the 2009 EQA, efforts were made to assess the quality of the

executive summary based on the observation that readers may read only this section when seeking to determine the usefulness of an evaluation report.



Data regarding missing executive summaries are summarized in the table below.

EQA Exercise	% Missing Executive Summary
2005	48%
2008	35%
2009	19%

Comparing the 2009 EQA results to those of 2005 and 2008, it is noteworthy that the percentage of evaluation reports missing an executive summary has declined to 19%, a substantial improvement since the 2005 EQA. Nevertheless, there appears to be substantial scope for improvement as 32% of the executive summaries were deemed to fall below expectations, leaving slightly less than half of all reviewed evaluations meeting expectations.

9.2) Completeness of Reports

ITEM	Proportion of Reports including Item		
	2005	2008	2009
Executive Summary	52%	65%	81%
Bibliography	48%	61%	49%
List of People met	54%	58%	54%
Data Collection Instrument(s)	32%	32%	41%
Terms of Reference	52%	68%	46%

The findings on completeness can be summarized as improving with regard to two of the five elements: inclusion of executive summary and data collection instruments and worsening for the other three: inclusion of bibliography, list of people met and terms of reference. The statement in the 2005 report – that including these items is an ‘easy fix’ - remains as applicable as when initially made.

CONCLUSIONS

- Previous EQA efforts attempted to estimate the number of evaluations done in and by UNFPA. In 2009, three rounds of solicitation by memo identified 57 evaluations from 123 country offices in the period 2007-2008, amounting to approximately 0.25 evaluations per CO per year. In practice, CO doing at least one evaluation appear generally more likely to do evaluations, since the 57 evaluations came from only 31 COs. Within the limits of the process used to identify evaluations for this EQA, fully 75% of UNFPA CO did no evaluation in 2007-2008. At this low level of evaluation coverage, UNFPA will likely be unable to identify lessons learned in any systematic way. Measures to record evaluations more systematically would provide some mitigation of this lost opportunity for learning and also enable Management to identify UNFPA business units who were doing too few (and too many) evaluations.
- Each of the findings has included summaries of performance and, where available, trends in performance. No systematic improvement in evaluation quality in the period 2007-2008 was identified. The persistently large proportion of evaluation reports that lack basic elements of evaluation, notably clear statements of evaluation questions, methods suitable to answer the evaluation questions, findings arising from data and evidence gathered through the application of said methods, and conclusions and recommendations grounded in the findings suggests that quality could be improved by implementing a systematic review and coaching process for UNFPA evaluations. Such a review process could involve RO and HQ resources or be done through peer networks among CO staff and would logically be supplemented by checklist-type tools available to UNFPA staff who design, commission and/or manage evaluations. Standardizing an ‘internal ToR’ to establish an expectation that these checklist tools be used before an evaluation is considered complete should reduce the persistently high proportions of missing items.
- Below expectation performance undermines the value of UNFPA’s resources invested in evaluation and also provides opportunities for improvement. For COs who use consultants to complete evaluations, steps to reduce the potential risk of principal-agent problems could improve evaluations. This problem may arise when a consultant’s primary interest is the maintenance of a business relationship with UNFPA rather than a delivering a high quality evaluation and is compounded when the CO staff lack the capacity or support to assess the quality

of the resulting evaluation work and/or focus on implementation (i.e. do an evaluation) rather than return on investment (i.e. do a good quality evaluation). Systems of peer or third-party review of ToR, inception reports and final reports combined with contracts that make final payment conditional on modification to address reviewers' concerns could assist in reducing the high proportion of UNFPA evaluations whose quality remains below minimum expectations..

RECOMMENDATIONS

- 1) The persistently high rates of missing items in evaluation reports is concerning and creates an opportunity to improve guidance to CO and implementing partners. Programme Division, with input from DOS, is recommended to develop and disseminate minimum expectations for evaluation reports to all CO, RO and HQ business units by no later than June 30, 2010. This effort should yield both checklist-type tools and the business process infrastructure needed to manage peer review and such other measures as Management implements to address the poor quality of UNFPA evaluations.
- 2) Particular attention is needed to country programme evaluations (CPE) to ensure that UNFPA complies with the requirement from decision 2009/18 that every programme be evaluated at some point in its cycle. Programme Division, supported by relevant direction from the OED, is recommended to communicate expectations regarding CPE to all CO whose programme cycles end in 2010 and 2011 by no later than March 31, 2010. In addition, including evaluation plans for CPE in draft country programme documents would enable evaluation design to mirror programme design. The Executive Committee is recommended to add coverage data for CPE to its review of the annual evaluation plan, beginning with the plan for the period 2010-2011.
- 3) Whether due to indifference to DOS requests for evaluation reports or the possibility that no evaluations are done in most UNFPA business units, current processes to identify and track evaluation reports require strengthening. The small number of evaluations apparently being done does not provide an adequate basis for quality assurance across UNFPA. Management is recommended to implement whatever measures, registry or otherwise, that will enable it to comply with EB decision 2009/18's requirement that every UNFPA programme be evaluated at least once in its programme cycle. These measures should be in place by June 30, 2010, prior to when DOS will request evaluations completed in 2009 for the 2010 EQA.
- 4) Pervasive weaknesses in assessment against the 5 OECD/DAC criteria are not a new EQA finding. Ensuring linkages between cost and expenditure information and programme activities, outputs and outcomes could enable a more consistent assessment of efficiency. The Programme Division, in its role leading results-based management, is recommended to convene a meeting, including field input,

to review how UNFPA might use available information to assess efficiency more consistently and develop a standard template for making such information available to contracted evaluators. This work should be completed by August 31, 2010.

- 5) The absence of documentation of ethical considerations is consistent with neither professional practice in evaluation nor the Fund's own guidelines. Programme Division, supported by the Ethics Office and DOS, is recommended to produce and disseminate a short summary of ethical considerations in evaluation and minimum documentation expectations to all UNFPA staff by June 30, 2010.
- 6) Consistent with the terms of decision 2009/18, DOS has planned to complete the EQA assessment annually to support the biennial reporting to the EB on evaluation. DOS has provided a guide to the EQA process on its Intranet site to encourage transparency and provide guidance to UNFPA staff who may be involved in managing or implementing evaluations. All CO staff commissioning, designing or implementing evaluations are encouraged to review this material.

ANNEX 1

EQA STANDARDS

DESCRIPTION: Objectives of the evaluation are clearly defined. Considerations have been given to ethics.

SPECIFICS:

1.1) Evaluation objectives 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.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Objectives are vague, and do not provide a clear statement of what the evaluation seeks	Objectives provide a clear statement of what the evaluation seeks to accomplish and evaluation questions are sufficiently detailed to clearly link to objectives and/or OECD-DAC criteria (efficiency, effectiveness, relevance, sustainability and impact)	Can be considered best practice among the sample of evaluation reports

1.2) The evaluation report describes mechanisms and measures that were implemented to ensure that the evaluation process conformed with relevant ethical standards including but not limited to informed consent of participants, privacy and confidentiality considerations.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Little (few sentences, scant detail) or no discussion of	Relevant ethical standards are identified and adequate details	Can be considered best practice among the sample of evaluation reports

	relevant ethical standards and how the evaluation addressed these	provided to provide reader assurance that standards were maintained	
--	---	---	--

DESCRIPTION: The scope of the evaluation is clearly defined and the design addresses relevant OECD/DAC criteria.

SPECIFICS:

2.1) The terms of reference (ToR) clearly describe the evaluation’s intended scope and focus –either by defining the main evaluation questions, and/or listing main objectives. The ToR specify evaluation criteria to be used given the evaluation’s objectives and scope, particularly the OECD/DAC criteria (relevance, efficiency, effectiveness, sustainability and impact).

NOTE: This item should be marked ‘missing’ if the evaluation report does not include ToR.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
ToR not included with evaluation report	Evaluation questions/objectives not or ill-defined and/or scope unclear	Scope well-described; questions/objectives clearly defined and relevant OECD/DAC criteria identified and consistent with scope and questions/objectives	Can be considered best practice among the sample of evaluation reports

2.2) Assessment of programme/project relevance examines the degree to which the outputs/outcomes are in line with national needs/priorities, UNFPA priorities, and relevant to stakeholders.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Evaluation assessed only one of linkage to national needs or UNFPA priorities	Evaluation assessed linkage to national needs, UNFPA priorities and relevance to stakeholders	Can be considered best practice among the sample of evaluation reports

2.3) Assessment of effectiveness examines the extent to which a programme/project achieves its planned results (outputs and outcomes).

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Evaluation assesses inputs only and/or does not include any causality analysis linking inputs to outputs/outcomes, including coverage if applicable	Evaluation assesses outputs and outcomes, achieved and planned and provides causality analysis linking inputs to outputs/outcomes, including coverage if applicable	Can be considered best practice among the sample of evaluation reports

2.4) Evaluation assesses efficiency, linking outputs to expenditures/resources and assessing whether this occurred as economically as possible.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Inputs and/or outputs identified but inputs not linked to outputs	Quantity and quality of inputs assessed and linked to outputs with some consideration of alternatives to ascertain how economical programme/project was	Can be considered best practice among the sample of evaluation reports

2.5) Sustainability assesses the extent to which programme/project results are likely to continue/remain after termination of external assistance

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Evaluation states results are sustainable (or not) but provides little/no evidence to support this statement	Evaluation states results are sustainable (or not) and provides evidence, notably capacity assessment to support this	Can be considered best practice among the sample of evaluation reports

		statement	
--	--	-----------	--

2.6) Impact assesses positive and negative long-term effects which may be economic, socio-cultural, institutional, environmental, technological or other effects.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	No assessment of impact or assessment of likelihood of long-term impacts	Impact or likelihood thereof assessed and supported by evidence, including causality analysis	Can be considered best practice among the sample of evaluation reports

DESCRIPTION: The methods are well-described, logically linked to evaluation questions, consistent with good practice, and include efforts to identify and control bias and/or acknowledge limitations.

SPECIFICS:

3.1a) Methods are valid i.e. focused on and logically linked to evaluation objectives and questions.

3.1b) Methods consistent with good practice and include, where appropriate explicit efforts to test counterfactuals and triangulate among data sources

3.1c) Efforts to control bias and/or acknowledge limitations due to uncontrolled bias are implemented and described

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Methods are poorly linked to evaluation questions/objectives OR methods fall short of good practice OR bias control efforts are inadequate	Methods are reasonably linked to evaluation question/objectives AND consistent with good practice, including counterfactual and/or triangulation as appropriate AND bias control addressed and described	Can be considered best practice among the sample of evaluation reports

DESCRIPTION: Stakeholder/Beneficiary Involvement

SPECIFICS

4.1) Consideration is given to stakeholder involvement and report is clear about rationale for and level of stakeholder involvement. Involvement may include participating in design (questions/objectives, methods, data collection instruments), collecting data, analyzing data, or developing recommendations, and other roles as appropriate for the evaluation under consideration.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Stakeholders not clearly defined OR treated as single group AND/OR extent of participation unclear	Stakeholders well-defined and details of participation of each stakeholder provided	Can be considered best practice among the sample of evaluation reports

DESCRIPTION: Gender Mainstreaming

SPECIFICS

5.1) Where relevant, evaluation assesses extent to which data disaggregated by gender were used for planning and assessing programme/project and extent to which programme/project promoted gender mainstreaming

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Gender mentioned but unable to determine if gender-disaggregated data used or not	Gender disaggregated data used to plan and assess programme/project; gender mainstreaming promoted	Can be considered best practice among the sample of evaluation reports

DESCRIPTION: Findings/Results follow logically from analysis, are credible and clearly presented together with analyses of achievements/deficiencies.

SPECIFICS

6.1) Findings/Results are analyzed in terms of outcomes or impacts, including cost analyses, and reasons for accomplishments/difficulties are identified and supported by analysis.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Vague or incomplete findings AND/OR little if any analysis of accomplishments and/or deficiencies	Findings clearly presented and grounded in analyses of data, including analyses of accomplishments and/or deficiencies	Can be considered best practice among the sample of evaluation reports

DESCRIPTION: Report presents clear conclusions supported by findings and analyses.

SPECIFICS

7.1) Conclusions flow logically from and are supported by evaluation findings and address issues of significance to the programme/project as scoped by the evaluation questions/objectives.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Conclusions do not follow from findings AND/OR not formulated in relation to evaluation questions/objectives	Conclusions are consistent with findings and supported by data analyses AND formulated in relation to evaluation questions/objectives	Can be considered best practice among the sample of evaluation reports

DESCRIPTION: Evaluation presents a set of useful, clear and practical recommendations

8.1) Recommendations are supported by data analyses, findings and conclusions, are clearly stated and include with specific details of who is recommended to do what by when. Where relevant, stakeholders have been involved in formulating recommendations.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Recommendations poorly linked to analyses, findings and conclusions OR	Recommendations grounded in analyses, findings and conclusions	Can be considered best practice among the sample of evaluation reports

	lack specific details of who does what by when	AND include details of who does what by when	
--	--	--	--

DESCRIPTION: Evaluation report is clear and well-structured, including an Executive Summary

SPECIFICS

9.1) Executive Summary is concise, readable and can stand alone without reference to the rest of the report.

Missing	Below Expectations (1-3)	Meets Expectations (4-6)	Exceeds Expectations (7-9)
	Executive Summary incompletely describes programme/project, major outcomes, conclusions or recommendations	Executive Summary is clear, readable and can stand alone	Can be considered best practice among the sample of evaluation reports

9.2) Completeness criteria (Y/N):

- Executive Summary (Y/N)
- ToR (Y/N)
- Bibliography (Y/N)
- Data Collections Instruments (or description thereof) (Y/N)
- List of people met (Y/N)