

UNICEF EVALUATION REPORT QUALITY REVIEW 2006

UNICEF's average evaluation report quality has increased considerably from 2000 to the present. In December 2006 the average rating of all reviewed 2005 evaluation reports was 2.89 (a Satisfactory rating leaning very much towards a Very Good) whereas in the year 2000 the average rating was only a 2.28, a Satisfactory close to a Poor rating.

Year	Number of Evals Reviewed	Average Rating
2000	78	2.28
2005	22	2.89
2006	6	3.65

All evaluation reports are rated against a set of 22 *UNICEF Evaluation Report Standards* that are drawn from and complementary to a broad range of widely accepted references on evaluation design and process. Desired outcome of the quality review process according to the *Standards* is an increased UNICEF evaluation report quality that will be expressed through high ratings (minimum Satisfactory) for all Standards.

Overall, twelve Standards now have a majority of evaluations receiving Satisfactory or better: Executive Summary; Programme description; Stakeholders' role and their contribution; Purpose and context; Evaluation criteria and use of OECD/DAC criteria; Objectives and Scope; Appropriate evaluation methods; Clear findings with inputs, outputs and outcomes measured; Reasons for accomplishments and difficulties; Substantial conclusions that provide insights in the problem; and Prioritized recommendations firmly based on evidence. By the end of 2006, UNICEF has 60% of the 20 Evaluation Standards with a majority of evaluations receiving Satisfactory or better – compared to only 50% reaching this goal in 2005.

Areas remaining for improvement are the description of stakeholder participation and analysis of relative contributions of stakeholders to the results, attention to eligible persons not reached by the project/programme, cost analysis, and a consideration of HRBAP, gender analysis, and RBM.

In 2005, standards #10, 11, 14, and 19 were introduced as key standards. Evaluations that fail (Missing or Poor rating) on the key standards are not posted on the Evaluation and Research Database (ERD). Standards #11 (appropriate evaluation methods) and #19 (conclusions) show considerable improvement over the years and are now met by more than 80% of all reports from 2005/06. Two key standards have yet to reach the Satisfactory threshold. Only 17% of all 2005/06 reports comply with Standard #14: A description of ethical considerations and safeguards is notably absent and may be a critical issue for validity of the reports to an outside reader. The other problem is in the transparent description of the methodology – 43% evaluations in 2005/06 still fail Standard #10. With respect to the key standards, a management response mechanism was introduced. COs with reports that fail a critical standard are contacted by the EO and encouraged to correct the problem.

Table of Contents

Background and Purpose	3
Objective	3
Methodology and Limitations	3
Availability of Reports Reviewed	6
Findings	6
Overall Results	6
Results for Individual Standards	9
Future Reviews	13
Conclusions/ Recommendations	13
Outlook 2007	14
List of UNICEF's Evaluation Report Standards	16
List of 2005/06 Evaluations that Received a 5 Excellent	17

BACKGROUND AND PURPOSE

1. In September 2004, the UNICEF Evaluation Office (EO) released *Evaluation Report Standards* to be included with every Terms of Reference for an evaluation. This document describes what UNICEF requires of every evaluation report commissioned by the organization in 22 concrete standards with detailed descriptions. The *Standards* also created a rating system to provide a transparent and objective monitoring and measurement tool regarding the quality of the evaluation report. The list of the 22 standards is provided at the end of this report.

2. This is the third annual review of the status of Country Office-commissioned evaluation reports using the *Standards*. The findings will provide preliminary results or indications for recent initiatives to improve the quality of Country Office commissioned evaluations and assist offices in adjusting those work plans. It will also serve as regular update to the Evaluation Committee.

OBJECTIVE

3. This review identifies areas of strength and weakness in general categories. Mainly, this review will talk about the evaluation reports as a whole. Since there were limited numbers of evaluations reviewed, findings are not analyzed on a regional basis.

METHODOLOGY AND LIMITATIONS

4. The criteria by which quality will be judged is the *UNICEF Evaluation Report Standards*. Each evaluation is given a rate on a five point scale for each of 22 standards and a weighted average of these ratings is computed for an overall rating. A complete description of the process is in the *Standards*.¹

Rating System	
1 – 1.99	Poor
2 – 2.99	Satisfactory
3 – 3.99	Very Good
4 – 5	Excellent, Model Evaluation

5. As of December 2006, UNICEF Country Offices have sponsored or conducted 1622 evaluations in the years 2000-2006. Each evaluation has been entered in the Evaluation and Research Database (ERD), where the submission status of each is noted. Submission status can be “submitted” (actual report sent to EO), “pending” (Executive Summary sent but report not yet sent), “not submitted” (no Executive Summary and no report sent). All reports that were “submitted”, “pending”, or “not yet submitted” are traditionally included in the Annual Review count to allow a consistent comparison with previous reviews.

6. It should be noted that further options to mark ERD entries exist. However, these do not refer to the submission status but to quality and implementation status of evaluations: “Not recommended” indicates that the CO does not recommend the evaluation because of low quality of the report. Low quality evaluations that are not approved by the CO did not have to be submitted (this feature will change in the future) and hence are not included in the quality review. “Planned” is a feature that was introduced in 2006 and provides an overview of planned reports for the coming year. “Cancelled” is an option to mark evaluations that - though announced - did not take place in the end. Evaluations marked according to these three options are not included in this or previous quality reviews.

7. The *Annual Quality Review* focuses on Country Office evaluation quality only since it is the COs that implement most evaluations (93% of all submissions to the ERD are from COs as of

¹ Typical DAC evaluation criteria of efficiency, effectiveness, impact, sustainability and relevance are not considered as they are appropriate to programme or project evaluations and not meta-evaluations of this type.

December 2006). In addition, only CO-led evaluations exist in sufficient regional and thematic diversity to permit the analysis featured in this report. Also, ultimately it is CO evaluation quality that reflects mostly if capacity building for evaluation achieved success. For purpose of comparison, however, it is useful to note that out of 47 evaluations from RO and HQ submitted to the ERD 25 reports have been rated with an average 3.00 (Very Good) for the years 2000-2006. If looked at over time the picture becomes even clearer with RO and HQ evaluations achieving an average rating of 2.7 (Satisfactory) from 2000-2003 and a 3.2 (Very Good) from 2004-2006.

Sample

8. Out of all 1622 ERD postings from 2000-2006, 601 reports (37%) were submitted to the ERD as of December 12, 2006. Out of the submitted evaluations, 313 reports were rated and also conducted by a Country Office and in English (up to 2006 the EO only had capacity to quality review English language reports). These evaluations constitute the sample for this review – representing 20% of all 1622 sponsored or conducted CO evaluations from the period 2000-2006.

Regional breakdown

9. A regional breakdown provides insights in submission activity by region. This table may be useful in particular for regions below the 40% submission threshold as an alert to intensify submission efforts and thereby contribute to institutional learning and knowledge sharing.

Region	Submitted	Pending	Not Submitted	% Submitted
CEE/CIS	88	21	121	38%
EAPR	75	4	136	35%
ESAR	127	30	151	41%
MENA	71	1	96	42%
ROSA	60	6	91	38%
TACR	85	46	175	28%
WCAR	95	9	134	40%
Total	601	117	904	37%

Evaluation postings (2000-2006) entered in the ERD as of December 12, 2006

Rating

10. On a five point scale, the ratings of the review sample range from Very Good 3.85 to Poor 1.26 with an average of 2.34. Ratings were simply totaled and a means taken.

11. A corporate “strength” is defined as a topic area or standard where at least 51% of the evaluations received Satisfactory or above (2 or greater.) Ideally, 100% of evaluations should meet this cut-off.

Previous reviews

12. This review makes comparison to the results of the 2005 *UNICEF Evaluation Report Quality Review* which had identical methodology. When data is presented from the previous review it will be shaded in gray.

Strengths

13. The Annual Review is based on a vast set of UNICEF and other international cooperation agency sources. A set of norms and standards (e.g. the *UNICEF Evaluation Report Standards* and the more recent UN Norms and Standards <http://www.uneval.org/docs/ACFFCA1.pdf>) is available for normative reference.

14. The Evaluation and Research Database provides for a carefully quality reviewed data set that allows for a sound and detailed analysis. It is a sample cross-section of UNICEF's evaluation work.

15. The criteria used in the quality review and rating process are the common UN norms.

16. Reviewers are professional staff. Further, when in doubt about a rating they get a second opinion.

17. Reviewers are trained on the standards and the rating process. They do not rate independently until they show consistency with existing raters.

Limitations

18. Compared to a potential 1622 reports sample, 313 reports may seem rather small. Two major limitations have been identified: There is an uneven regional distribution of reviewed reports due to language constraints (which has also influenced submission rates). This led to an over sampling of regions where English is the dominant language and to an unavoidable bias to the English languages reports in regions with many non English reports. This obstacle will be overcome in the 2007 review. Two consultants were recruited in 2006 to quality review all Spanish and French language evaluations (~150) accumulated since 2000. Their contributions will be included in the next review.

19. The second limitation is low submission rates. Only 37% of all evaluation reports 2000-2006 are actually submitted. In 2005, only 27% of reported evaluations were received compared to 33% in 2004 and up to 52% in 2000. For 2006, the total number of evaluations conducted is still unknown since the Annual Reports for 2006 have not yet been received but 361 evaluations have been announced as planned for 2006. Up to now, 10 reports have already been received.

20. A reason for low submission rates may be that most of available reports have already been submitted to the database and others may not be easy to find. Changing CO staff may further slow down easy retrieving of older reports. Another reason could be that reports received by the Evaluation Office are those that the Country Office believes are better and others are "not recommended" for submission – a kind of self-selection by COs. For 2007, COs will be required to submit also 'not recommended' evaluations. It is hoped to increase submission rates and subsequently generate a more comprehensive picture of evaluations conducted in UNICEF.

21. Another, though minor, restricting factor is a backlog of 40 reports in line for quality review. The backlog has to be ascribed to a change in 2006 of the person in charge of the quality review, to a higher volume of reports from COs and HQ in 2006 compared to previous years, and to time given to the recruitment process and training for Spanish/French quality reviewing consultants.

22. Findings in this review were not cross-checked by another source of data such as interviews or other reports. The evaluation report themselves were the sole source of information. It should be noted that this review is only an indirect assessment of evaluation quality. The Evaluation Office recognizes that the evaluation process with the participation and involvement of stakeholders can have significant value that may not be captured by a review of the evaluation report alone. However it is critical that the evaluation report be an accurate reflection to increase information sharing between offices and also externally with our partners and the public.

AVAILABILITY OF REPORTS REVIEWED

23. For evaluations that are considered Satisfactory or better and meet key standard #19 regarding conclusions, the full report and rating with comments can be found on the [Evaluation and Research Database](#). This is accessible to all staff via the Intranet.

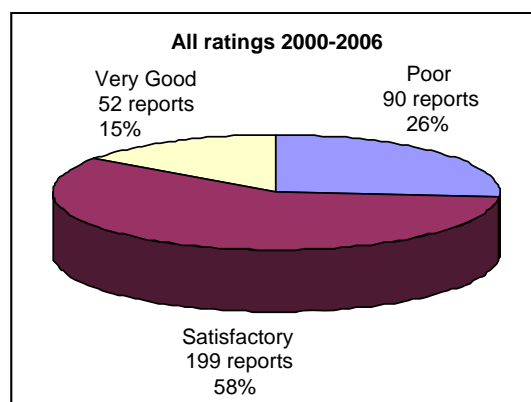
24. For evaluations that are considered Poor or fail to meet key standard #19, access is controlled. For Regional and Headquarters offices, all of the reports reviewed are available on the [Evaluation Report Submission Website](#). For Country Offices, only their own Poor reports can be viewed on this Website with a complete rating and comments describing where the report needs improvement. Country Offices cannot see the Poor reports of other Country Offices.

25. All satisfactory or better rated evaluations are not only posted on the ERD but also on the internet ([UNICEF - Evaluation database - Introduction](#)) and accessible for the public – with the only exception of reports that are of extreme sensitivity or include confidential issues that are only accessible upon request. UNICEF evaluations are also actively shared with other UNEG members through the website of the UN Evaluation Group (UNEG).

FINDINGS

OVERALL RESULTS

In total, the ratings for CO, RO, HQ evaluations range from Very Good (3.81) to Poor (1.26) with an average of 2.34 from 2000-2006. The majority (58%) of reports were in the Satisfactory range while 15% were considered Very Good and 26% considered Poor.



Trend analysis

26. Looking at the trend of ratings over the time from 2000, a remarkable increase in evaluations with a Very Good rating from single percentages (2000-2003) to 59% in 2005 is to be noted. At the same time, a reduction in the Poor rated reports takes place. For 2006 – not included due to limited sample size – six out of ten submitted evaluations were already rated Very Good. This increase in evaluation quality correlates with the introduction of the UNICEF evaluation report standards in 2004.

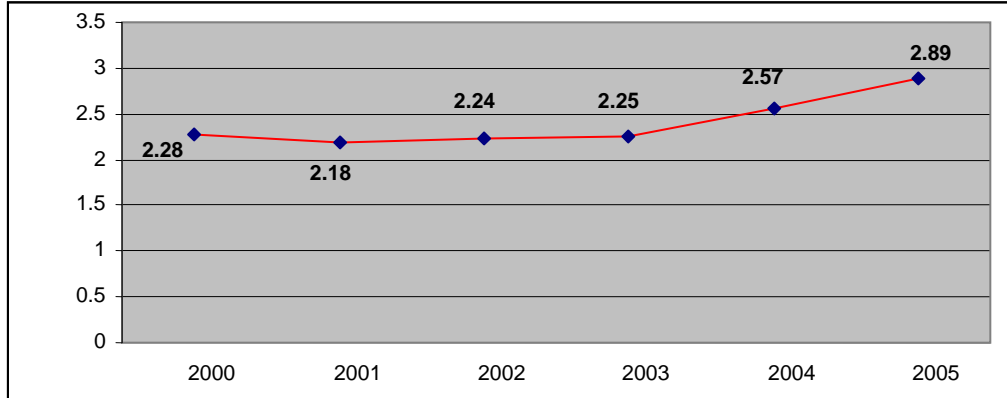
27. Not only did the share of Very Good evaluations rise and Poor ratings decrease. As illustrated below the quality of evaluation reports also rose in terms of average rating from 2.28 in 2000 to a 2.89 rating in 2005, thereby driving from a rather low range Satisfactory towards a highly Satisfactory quality.

Evaluation report (CO, RO, HQ) quality, 2000–2005

	2000	2001	2002	2003	2004	2005
Very Good (3 – 3.99)*	6%	4%	3%	6%	25%	59%
Satisfactory (2 – 2.99)	68%	56%	64%	58%	64%	28%
Poor (1 – 1.99)	26%	40%	33%	35%	11%	13%
Average Rating	2.28	2.18	2.24	2.25	2.57	2.89

* The rating 'Excellent (4 – 4.99)' has not yet been given to any evaluation and is therefore not listed.

Average rating by year, 2000-2005



Regional breakdown

28. A regional breakdown comparing this review to last years reviews shows improvement across the board. The small sample size for some regions should be noted. TACR and WCAR are especially limited since most of their evaluation reports are in Spanish and French respectively. Ratings for both regions will be available in the next year as earlier described.

2004 - 2006	Avrg Rating	# of Evals	Highest Rating	Lowest Rating	# Very Good	% Very Good	# Satisfactory	% Satisfactory	# Poor	% Poor
CEE/CIS	2.53	18	3.89	1.61	8	44%	5	28%	5	28%
EAPR	2.79	9	3.62	2.08	3	33%	6	66%	0	0%
ESAR	2.48	35	3.67	1.66	6	17%	24	69%	4	11%
MENA	2.79	13	3.31	1.97	6	46%	6	46%	1	8%
ROSA	3	10	3.59	1.84	5	50%	4	40%	1	10%
TACR	2.63	1					1			
WCAR	2.9	3	3.58	2.07	2	66%	1	33%	0	0%
2000 - 2003	Avrg Rating	# of Evals	Highest Rating	Lowest Rating	# Very Good	% Very Good	# Satisfactory	% Satisfactory	# Poor	% Poor
CEE/CIS	2.03	45	3.25	1.26	1	22%	19	42%	25	56%
EAPR	2.25	46	3.56	1.48	4	9%	29	63%	13	28%
ESAR	2.43	54	3.41	1.68	3	6%	42	78%	9	17%
MENA	2	32	2.95	1.56	0	0%	14	44%	18	56%
ROSA	2.27	31	3.36	1.39	2	6%	22	71%	7	23%
TACR	1.98	18	2.62	1.64	0	0%	11	61%	7	39%
WCAR	2.4	10	2.87	1.95	0	0%	9	90%	1	10%

29. In addition to average rating, it is worth looking at how many evaluations achieved Excellent (5) on one standard, which indicates that it is a model evaluation for that particular standard. In ROSA, 6 out of 10 evaluations from 2004-2006 were excellent in some aspect – a remarkable achievement.

2004 – 2006	# of Evals w a 5	% of Evals with a 5	# of 5s	2000 - 2003	# of Evals w a 5	% of Evals with a 5	# of 5s
CEE/CIS	7	39%	15	CEE/CIS	3	67%	5
EAPR	3	33%	7	EAPR	8	17%	12
ESAR	8	23%	10	ESAR	10	19%	10
MENA	5	38%	9	MENA	0	-	0
ROSA	6	60%	11	ROSA	5	16%	5
TACR	0	-	0	TACR	1	6%	1
WCAR	2	67%	3	WCAR	1	10%	2
Total	31	35%	55	Total	28	12%	35

Thematic breakdown

30. In the following discussion, please keep in mind the very small sample sizes under consideration: The comments are suggestive but cannot be considered definitive.

31. Looking at the Themes of the evaluations rated, four areas are strong with all of the evaluations in that topical area being of Satisfactory quality or better in the period 2004-2006: Programme Review (which only includes Country Programme Evaluations that were not lead by NYHQ), Capacity building, Gender Equity, and Nutrition had not a single Poor rating in 2004-2006. However, for the last three topic areas only one evaluation has been rated and hence represents a too small sample to allow for strong conclusions.

32. Emergency, WES, and ECD show a remarkable decrease of Poor rated evaluations. Only 5% evaluations in Emergency were rated poor compared to 20% in the period 2000-2003. In WES, the quality improved even from 52% poor to only 11%. ECD reports, till 2003 still of concern with 44% poor, sunk to a 20% level.

33. Areas of concern in the first three years of ERD introduction (2000-2003) were Advocacy and Communication, WES, and Child Protection. No Advocacy and Communication evaluation has been submitted in 2004-2006. WES and Child Protection experienced a shift and improved from 52/47% to only 11/13% respectively of evaluations rated Poor.

34. A corporate “strength” is defined as a topic area or standard where at least 51% of the evaluations received Satisfactory or above. Ideally, 100% of evaluations should meet this minimum cut-off. In the charts below, a strength will be represented by **% of Poor + Missing below 50%**.

2004- 2006	# of Evals	# of Evals Rated Very Good	# of Evals Rated Satisfactory	# of Evals Rated Poor	% of Evals rated Poor
Advocacy and Communication	0				
Capacity Building	1	1			0%
Child Protection	8	2	5	1	13%
ECD	5	2	2	1	20%

Education Girls	5	2	1	2	40%
Education Other	11	2	6	3	27%
Emergency	19	11	7	1	5%
Gender Equity	1		1		0%
Health	2		1	1	50%
HIV/AIDS	11	1	8	2	18%
Immunization	2		1	1	50%
Income Generation					
Nutrition	1		1		0%
WES	9	2	6	1	11%
Programme Review	9	4	5		0%
Total	84	27	44	13	15%

Review of 2000 – 2003

Theme	# of Evals	# of Evals Rated Very Good	# of Evals Rated Satisfactory	# of Evals Rated Poor	% of Evals rated Poor
Advocacy and Communication	4	1	1	2	50%
Capacity Building	4		4		0%
Child Protection	15		8	7	47%
ECD	18		10	8	44%
Education Girls	13	1	9	3	23%
Education Other	35		22	13	37%
Emergency	25	2	18	5	20%
Gender Equity	6		5	1	17%
Health	24	1	17	6	25%
HIV/AIDS	20		14	6	30%
Immunization	8		6	2	25%
Income Generation	4		3	1	25%
Nutrition	9	1	4	4	44%
WES	21	2	8	11	52%
Programme Review	8	1	5	2	25%
Total	214	9	134	71	33%

RESULTS FOR INDIVIDUAL STANDARDS

35. For the entire period 2000-2006, out of 20² standards eight can be considered strengths with a majority of the evaluations receiving Satisfactory or better. A look at the more recent and shorter period 2004-2006 shows that four additional standards reached the threshold of more than 50% plus one standard at exactly 50%. If considering only reports from 2004 on, UNICEF could name twelve areas as corporate strengths. This indicates that more recent reports received better ratings for standards #2,6,21, and 4 and points towards a trend of higher evaluation quality in the last years.

² Two standards - #1 and #22 – are not included in the analysis for this section because they represent simple check lists concerning the title page and annexes.

36. With respect to the single standard results it should be noted that 'no information' equals a Missing rating. This has an important impact when looking at the aggregated standard results. For example, the highest Poor+Missing rating of standard #14 is mainly based on the fact that no information on this standard was provided but in no case signifies that 91% of all evaluations were unethical.

Standard	% Poor + Missing	
	2000-2006	2004-06
3. The project/programme to be evaluated was clearly described, including the logic of the programme design or expected results chain.	29%	19%
19. Conclusions were substantiated by findings consistent with data and methods and represent insights into identification and/or solutions of important problems or issues.	36%	23%
7. Objectives were realistic and achievable in light of the information that can be collected in the context of the programme/project. The scope of the evaluation was defined clearly.	32%	25%
11. Evaluation methods were appropriate and adequate providing a complete and fair assessment. Consideration was given to limitations of the methodology.	41%	25%
15. In presenting the findings, inputs, outputs, and, where possible, outcomes/impacts were measured (or an appropriate rationale given why not).	44%	28%
18. Reasons for accomplishments and difficulties, especially continuing constraints, were identified as much as possible.	45%	30%
20. Recommendations were firmly based on evidence and analysis; they were directly relevant and realistic with priorities for action made clear.	36%	33%
5. Purpose and context provided a specific justification for undertaking this evaluation at this time.	48%	33%
2. Assessment of Executive Summary	52%	36%
6. The evaluation used standard OECD/DAC evaluation criteria as per UNICEF PPPM and/or provided an explanation for criteria that was considered not applicable and not used.	54%	42%
21*. <i>Lessons learned, when presented, were generalized beyond the immediate intervention being evaluated to indicate what wider relevance there might be.</i>	51%	45%
4. The role and contributions of UNICEF and other stakeholders to the project/programme were clearly described.	56%	45%
9. The evaluation design considered programme/projects' incorporation of results based management.	68%	50%
10. Transparent description of methodology.	69%	50%
8. The evaluation design considered programme/projects' incorporation of the UN and UNICEF's commitment to human rights-based approach to programming.	80%	53%
17. Analysis of results includes appropriate discussion of the relative contributions of stakeholders to results.	65%	53%
12. A complete description for stakeholder participation was given.	85%	64%
16. To the extent feasible, the report includes cost analysis.	83%	67%

13*. Where information was gathered from those who benefited from the programme/project, information was also gathered from eligible persons not reached.	92%	82%
14. The evaluation design was ethical and included ethical safeguards where appropriate, including protection of the confidentiality, dignity, rights and welfare of human subjects, particularly children, and respect for the values of the beneficiary community.	98%	91%

**Standard 13 and 21 are the only standards that can receive a “0” – for not applicable. Because of this particularity not all reports can be considered and the baseline for the calculation of the indicator has been adjusted for reports only that received a rating other than “0”.*

Analysis of the Key Standards 10, 11, 14, 19

37. The Evaluation Office began to implement the key methodology standards in 2005. Key standards are #10, 11, 14, and 19. If the report receives a 'Poor' or 'Missing' on a key standard, the report will not be posted on the Evaluation and Research Database (ERD), regardless of the overall rating received. For evaluations completed in 2005, letters will be sent to Country Offices when a report does not meet a key standard. Evaluations completed in 2006 that do not meet a key standard will not be included in the ERD. This is not to say that the evaluation was not methodologically sound and/or ethical but rather that the report lacks information on this topic for an outside reader to be assured of its propriety and validity. For reports before 2006, the CO is invited to correct the problem by sending an amendment to the EO or having the consultant correcting the problem in the report – with this the revised report then qualifies to be included in the ERD. Of the nine COs contacted, two have been able to eliminate the problem.

10. Transparent description of methodology						
Standard 10	2000	2001	2002	2003	2004	2005-2006
# Excellent	0	0	0	0	1	3
# Very Good	0	0	1	2	6	7
# Satisfactory	8	8	15	7	12	7
# Poor	41	22	23	19	34	13
# Missing	4	3	1	1	2	0
% Poor + Missing	85%	76%	60%	69%	65%	43%

11. Evaluation methods were appropriate and adequate providing a complete and fair assessment. Consideration was given to limitations of the methodology.						
Standard 11	2000	2001	2002	2003	2004	2005-2006
# Excellent	0	0	0	0	1	0
# Very Good	2	0	3	3	14	15
# Satisfactory	17	19	20	13	21	10
# Poor	31	12	17	10	18	5
# Missing	3	2	0	3	1	0
% Poor + Missing	64%	42%	43%	45%	35%	17%

35. Standard #10 asks if the description of the methodology is complete. Standard 11 asks if the methodology itself is appropriate and adequate. There is usually some leeway on Standard 11 when standard #10 is poor which may explain the difference in ratings. Although poorly described, the methodology may indeed be valid.

36. This should not give the impression that methodology description is not important. Both of these standards are considered key standards. In addition to risking the credibility of the report, it loses some of its value as a learning tool when methodology is not clear. Others cannot learn and replicate the evaluation framework if it is not described well.

37. Both standards improved considerably in 2005-06 with far fewer 'Poor' or 'Missing' ratings.

14. The evaluation design was ethical and included ethical safeguards where appropriate, including protection of the confidentiality, dignity, rights and welfare of human subjects, particularly children, and respect for the values of the beneficiary community.						
Standard 14	2000	2001	2002	2003	2004	2005-2006*
# Excellent	0	0	0	0	0	0
# Very Good	0	0	0	0	0	0
# Satisfactory	2	0	0	0	2	5
# Poor	0	1	1	1	2	1
# Missing	51	32	39	28	51	24
% Poor + Missing	96%	100%	100%	100%	96%	83%

38. Standard #14 continues to be a problematic area. It is the standard with the highest percentage Missing or Poor with an average of 98% for the entire period 2000-06 and still 83% for last year's reports. Evaluation teams should note that a concise paragraph is usually all that is needed, however, the paragraph is essential to ensure the reader of the propriety of the evaluation.

19. Conclusions were substantiated by findings consistent with data and methods and represent insights into identification and/or solutions of important problems or issues.						
Standard 19	2000	2001	2002	2003	2004	2005-2006
# Excellent	0	0	0	0	2	1
# Very Good	5	5	7	6	14	11
# Satisfactory	29	14	18	11	22	13
# Poor	15	11	15	10	16	4
# Missing	4	3	0	2	1	1
% Poor + Missing	36%	42%	38%	41%	31%	17%

39. Standard #19 Conclusions is one of the overall strengths of UNICEF evaluations. The feature that prevents evaluations from moving from Satisfactory to Very Good is the ability to go beyond a simple presentation of results consistent with data to conclusions that represent insights into identification of problems or issues.

FUTURE REVIEWS

40. As an annual exercise, this review will be expanded in the next issue, December 2007. For the first time, Spanish and French language reports will be reported highlighting those regions with a majority of Spanish and French language reports, Latin America and West-Africa. It should be noted that results for the two regions will be available earlier than next December 2007. Readers particularly interested in French and Spanish evaluation quality can contact the EO from February 2007 (aroth@unicef.org). Also, individual report ratings will be posted on the ERD and fed back to the COs as ratings are available.

CONCLUSIONS/ RECOMMENDATIONS

41. This review provides preliminary findings that the quality of CO sponsored or conducted evaluation reports is improving. The increase in average rating as well as higher rated individual standards show that the system is responding to stimuli provided.

42. The improvement in report quality can be attributed largely to the combination of the introduction of the evaluation report standards and activities such as more advocacy about

evaluation quality by the EO, better hiring of M&E staff, and the issuance of the TOR technical note. However, a critical role remains with the Evaluation and Research Database's inherent quality review mechanism as it is clear that the knowledge that the reports will be rated causes COs to be more attentive to quality. The development since introduction of the ERD raises hope for a soon to be reached Very Good or even Excellent quality of all UNICEF evaluations.

43. The gap between satisfactory and excellent evaluation quality may be filled by ensuring that the upward trend of previous years will be continued. To increase further quality and not to forget also quantity of accessible evaluations some steps can be taken:

Quality of CO evaluations

44. COs should ensure that all necessary description is included in the evaluation report with special focus on

- Evaluation methodology with a description of ethical considerations and safeguards
- HRBAP, gender analysis/data disaggregation and RBM
- Description and address of the role and contributions of partners, including their role in the evaluation itself
- Key standards
- Comprehensiveness of information and format

45. COs are to use the *UNICEF Evaluation Report Standards* and make sure the evaluation team/consultant are provided with a copy during the TOR design and recruitment process.

46. Training on the Standards can be provided for COs/ROs and questions are always welcome (contact aroth@unicef.org)

Quantity

47. Organizational learning relies on the increase of evaluation submission rates. Having a representative sample at hand is not only important for learning from successes and failures of CO project and programmes but also to create an evidence-base for UNICEF activities and strategies. COs should pay special attention to share information and make sure that reports are sent as soon as finalized to NYHQ using the Evaluation Report Submission Website for everyone to access. If there is problem with submitting reports the EO should be contacted (Isanagustin@unicef.org).

OUTLOOK 2007

Evaluation and Research Database (ERD) and Knowledge Management

48. The ERD is widely recognized as a high quality example of a UNICEF knowledge management tool. In KM working groups and in bilateral discussions EO-PD sections it is currently discussed how to promote the contents of the ERD fundwide and intensify its use by PD sections. First steps have already taken, e.g. a pilot initiative with Nutrition section. To serve better the organization, the ERD is constantly improved according to user needs. Recommendations and comments on the ERD are always encouraged.

Studies and Surveys

49. To the moment, not only evaluations but also research reports (studies and surveys) are contained in the ERD. This feature was included in the ERD since there was no other fund-wide repository for research reports. At present, 602 surveys and 939 studies are housed. Till today, the ERD remains the only tool that systematically files studies and surveys. Whereas evaluation reports clearly belong under the umbrella of the EO, research reports remain waiting for one or more 'owners' who are able to process the vast knowledge contained in them. Some PD sections showed interest in developing a strategy on how to use studies and surveys of their area in 2006. For 2007, ownership of studies and surveys should be a topic on the list to resolve.

'Not recommended' reports

50. For 2007, some actions to improve the ERD are planned. As already announced in para 20 from 2007 on 'Not recommended' evaluations will have to be submitted despite of not being recommended. With this, the EO hopes to prevent COs to set too high quality levels and maybe to keep back evaluations that contain valuable information but may not comply with some standards. Reports that do not comply with some standards still can greatly help organizational learning. It should be kept in mind that the ERD is primarily an organizational learning tool and can only provide that much information that has been fed into.

ERD Taxonomy

51. It has been noted that the current taxonomy of the ERD is not aligned with the most recent programme taxonomy of the MTSP and the PIDB. For 2007, the ERD taxonomy will be harmonized with the PIDB taxonomy.

Other actions

52. Other actions planned for 2007 include increased efforts to promote the ERD organization wide as knowledge management tool, to disseminate results of outstanding evaluations via media notes, and pilot-test lessons learned write-ups. With all the actions recommended and planned the next Annual Review may include reporting on even better evaluation quality and hopeful base its analysis on a larger sample thanks to increased submission rates.

List of UNICEF's Evaluation Report Standards

1. Completeness of Title Page and Opening Pages

Executive Summary:

2. Assessment of Executive Summary

Background and Purpose:

3. The project/programme to be evaluated was clearly described, including the logic of the programme design or expected results chain.
4. The role and contributions of UNICEF and other stakeholders to the project/programme were clearly described.
5. Purpose and context provided a specific justification for undertaking this evaluation at this time.

Objectives and Evaluation Design:

6. The evaluation used standard OECD/DAC evaluation criteria as per UNICEF PPPM and/or provided an explanation for criteria that was considered not applicable and not used.
7. Objectives were realistic and achievable in light of the information that can be collected in the context of the programme/project. The scope of the evaluation was defined clearly.
8. The evaluation design considered programme/projects' incorporation of the UN and UNICEF's commitment to human rights-based approach to programming.
9. The evaluation design considered programme/projects' incorporation of results based management.

Methodology:

10. Transparent description of methodology
11. Evaluation methods were appropriate and adequate providing a complete and fair assessment. Consideration was given to limitations of the methodology.
12. A complete description for stakeholder participation was given.
13. Where information was gathered from those who benefited from the programme/project, information was also gathered from eligible persons not reached.
14. The evaluation design was ethical and included ethical safeguards where appropriate, including protection of the confidentiality, dignity, rights and welfare of human subjects, particularly children, and respect for the values of the beneficiary community.

Findings:

15. In presenting the findings, inputs, outputs, and, where possible, outcomes/impacts were measured (or an appropriate rationale given why not).
16. To the extent feasible, the report includes cost analysis.
17. Analysis of results includes appropriate discussion of the relative contributions of stakeholders to results.
18. Reasons for accomplishments and difficulties, especially continuing constraints, were identified as much as possible.
19. Conclusions were substantiated by findings consistent with data and methods and represent insights into identification and/or solutions of important problems or issues.

Recommendations and Lessons Learned:

20. Recommendations were firmly based on evidence and analysis; they were directly relevant and realistic with priorities for action made clear.
21. Lessons learned, when presented, were generalized beyond the immediate intervention being evaluated to indicate what wider relevance there might be.

22. Completeness of Annexes

List of (CO, RO, HQ) 2005/06 Evaluations that Received a 5 Excellent

Country	Title	Rating	Year	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21
Georgia	Evaluation of the Family Support & Foster Care Project and Prevention of Infant Abandonment and De-institutionalisation Project	3.85	2006	5										5									
Serbia and Montenegro	External Evaluation of the Day Care Center "TISA" for Children with Disabilities, Bijelo Polje, Montenegro	3.05	2005			5													5				
Macedonia	Evaluation of Young People's Participation. Report from the Former Yugoslav Republic of Macedonia	3.36	2005	5			5							5							5		
Cambodia	Country Programme Evaluation Royal Government of Cambodia and UNICEF 2001-2005	3.62	2005	5					5	5											5		
EAPRO Regional	UNICEF Tsunami Communications Evaluation	2.43	2005																		5		
Tanzania	Evaluation of Tahea Supported 'Mama Makubwa' Initiative in Makete and Iringa Regions	2.9	2005																			5	
Kenya	End of Project Evaluation on the ECHO funded GOK/UNICEF Response to the Kenya Drought Emergency 2004 -2005	3	2005	5																			
Zambia	Promotion of Water, Sanitation and Hygiene Education in Choma, Namwala, Gwembe and Siavonga Districts	3	2005											5									
Zambia	Report on the Evaluation of the Child Rights Clubs Project in Zambia	3.67	2005					5				5											
Yemen	Assessment of the Child Development Project	3.11	2005		5	5																	
MENA Regional	Evaluation of UNICEF responses to recent earthquakes in Algeria, Iran, Morocco	3.31	2005			5														5	5		
Nepal	Participatory Assessment of the School Sanitation and Hygiene Education (SSHE) Programme in Nepal	3.28	2005									5											
Afghanistan	Country Programme Evaluation Government of the Transitional Islamic State of Afghanistan/UNICEF (2003 - 2005)	3.46	2005				5		5														
Global	<i>The 2004 Indian Ocean Tsunami Disaster Evaluation of UNICEF's Response (Emergency and Initial Recovery Phase)</i>	3.54	2005				5	5			5												
Global	<i>Evaluation of DFID-UNICEF Programme of Cooperation to Strengthen UNICEF Programming as it Applies to Humanitarian Response, 2000-2005</i>	3.63	2005			5														5	5		
Global	<i>Evaluation of UNICEF's support to mine action</i>	3.56	2006							5											5		

<i>Global</i>	<i>UNICEF's use of the United Nations Trust Fund for Human Security (2000-2004) – An Internal Assessment</i>	<i>3.02</i>	<i>2005</i>							<i>5</i>											
<i>Global</i>	<i>Evaluation of UNICEF Learning Strategy to Strengthen Staff Competencies for Humanitarian Response 2000-2004</i>	<i>3.03</i>	<i>2005</i>					<i>5</i>		<i>5</i>											