

GEROS – Global Meta-Evaluation Report 2012

Final Report

By:
Universalial Management Group



August 2013



Executive Summary

The Global Evaluation Reports Oversight System (GEROS) is a UNICEF-wide system managed by the Evaluation Office. It is used to assess the quality of final evaluation reports.

The main objectives of GEROS are: to provide senior managers with an independent assessment of the quality of their evaluation reports; to strengthen the internal evaluation capacity to improve the quality of the evaluations; to contribute to corporate knowledge management and organizational learning; and to report to Senior Management on the quality of evaluation reports – generating key performance indicators that point to the improvements, strengths, and weaknesses of overall final evaluation reports. To meet these objectives, all final evaluation reports are assessed and rated externally by an independent firm according to the UNEG Evaluation Report Standards for UNICEF, and a final analysis is done using all of the reviewed evaluation reports’ ratings to study their overall trends, strengths and weaknesses.

Methodology

The quality review process and this meta-evaluation cover all of the 2012 reports submitted to the UNICEF Global Evaluation Database before May 2013 (a total of 79 evaluation reports). Each evaluation was reviewed using a tool that is based on the UNEG Evaluation Report Standards for UNICEF.¹ This tool contains a total of 6 sections and 22 sub-sections, which comprise a total of 58 guiding questions. Each section in the review tool was rated according to a four-point performance scale: “Outstanding, Best Practice,” “Highly Satisfactory,” “Mostly Satisfactory,” and “Unsatisfactory,” as shown below. Reports were each given an overall rating based on the same scale. Individual questions were rated according to the first row of the scale below:

Colour Coding	CC	Dark green	Green	Amber	Red	White
	Questions	Outstanding	Yes	Mostly Satisfactory	No	Not Applicable
	Section & Overall Rating	Outstanding, best practice	Highly Satisfactory	Mostly Satisfactory	Unsatisfactory	

In addition, reviewers provided comments to justify each rating and constructive feedback for future reports. The individual reviews were subject to a quality assurance peer-review process to ensure consistency in the ratings.

The meta-analysis in this report is based on the aggregation of all the ratings and comments in the 79 reports reviewed, which served to generate data on overall trends, as well as strengths and weaknesses of the evaluation reports reviewed. Comparisons with previous years are provided, where data is available. A combination of qualitative and quantitative methods is used to examine each of the variables.

Findings

Report Quality – Overall and by Region

While 86 reports were reviewed for 2011, a total of 79 reports were assessed for 2012. The regions that submitted the largest number of evaluation reports were Eastern and Southern Africa Regional Office (ESARO) with 20 reports, West and Central Africa Regional Office (WCARO) with 19, Central and Eastern Europe, Commonwealth of Independent States (CEECIS) with 13, and HQ with 10. The Middle

¹ The review tool is available in Appendix V.

East and North Africa (MENARO) and East Asia and Pacific Regional Offices (EAPRO) only submitted 3 evaluation reports. This represents a significant decrease for EAPRO, having submitted 11 reports in 2011. The South Asia (ROSA) and Americas and Caribbean (TACRO) regional offices also submitted relatively few reports (6 and 5, respectively).

The quality of reports submitted to UNICEF increased dramatically between 2011 and 2012. The number of good quality reports – rated either outstanding/best practice or highly satisfactory – increased by 20% (from 42% to 62%). In 2012, only 4 reports (8%) were considered unsatisfactory, and a greater number of reports were rated mostly satisfactory (30%). For the most part, these ratings were attributed to reports that failed to address a large portion of the required Geros criteria and questions. Although no direct correlation can be established between the inclusion of the ToRs and the quality of a report, review data shows that reports with annexed ToRs tended to receive a higher rating (70% were rated as highly satisfactory or outstanding, while only 48% of reports without ToRs received similar ratings).

Overall Ratings for 2012 Reports



With a few exceptions, the quality of UNICEF-sponsored evaluations has also improved across Regional Offices, and the proportion of unsatisfactory reports diminished across all regions. Indeed, the number of good quality reports from WCARO and CEECIS increased by 28% and 37% respectively. Similarly, the quality of HQ EO-led and HQ Corporate evaluations improved greatly, with the number of good reports increasing from 57% in 2011 to 90% in 2012. TACRO and EAPRO fared poorly this year, with only 33% and 20% of reports rated as good quality. It must be noted that only a very small number of reviews were submitted for these regions in 2012, which meant that a few poor quality reports had a disproportionate effect on their overall ratings.

Report Classifications

In terms of **geographic scope**, 86% of evaluations reviewed for 2012 were national and sub-national. For both categories, quality has been steadily increasing since 2010.

With respect to **management**, almost half of the reviewed evaluations (48%) were overseen by UNICEF, and a few were led by country governments. The ratings for evaluations managed or co-managed by UNICEF tended to fare better, and their quality has improved markedly since 2011.

The range of evaluation **purposes** addressed in 2012 differed only slightly from previous years. The classification of reports by purpose (programme, policy, humanitarian, country evaluations, etc.) shows that programme evaluation continues to be the most common type of evaluation conducted (41% in 2012). The quality of reports has increased for nearly all categories of evaluation purposes.

The number of evaluations assessing **results** at outcome and impact has fluctuated over time. Reports that focus on output-level results, which are fewer in number and often constitute humanitarian or real-time evaluations, demonstrate the greatest improvement in terms of quality. Reports that address outcomes continue to under-perform with regard to meeting UNICEF quality standards.

No significant trends can be derived from an analysis of programme or project **stages** (i.e. summative, formative, or both). Most reports are summative or present a mix of summative and formative elements.

A large proportion of evaluations cover more than two of the **MTSP focus areas** (known as multi-sector evaluations). Since 2011, performance levels have increased in most of the focus areas. Policy Advocacy and Partnerships as well as HIV/AIDS and Children received the lowest ratings, but also include fewer evaluation reports.

In terms of report **independence**, the move towards independent external evaluations noted in the last meta-evaluation was maintained and even increased in 2012.

Report Content

The Geros assessment is structured around six categories based on the UNICEF-adapted UNEG Standards for Evaluation reports. In 2012, quality increased across all of these categories. Reports for 2012 were particularly strong in articulating the underlying context of the intervention being evaluated; describing the object of the evaluation; articulating the purpose, objectives and scope; explaining the methodological choices made; using the evidence objectively to construct findings; connecting results with the causal reasons for success or failure; and formulating relevant and clear recommendations.

However, some of the common problems encountered included: the absence of a theory of change or results logic; failure to specify the underlying questions and criteria guiding the evaluation; providing an insufficient description of ethical considerations taken into account in the evaluation process; the absence of a cost analysis; and conclusions that did not add value to the report findings.

The integration of the Human Rights-Based Approach to Programming (HRBAP), gender equality and equity dimensions continues to be generally weak in evaluations. The ratio of reports that successfully integrated these elements did not rise above 50%, such that gender was addressed in 46% of reports, 44% paid attention to HRBAP, and 41% incorporated a greater focus on equity. However, in line with previous years, the trend towards increased integration of human rights, gender and equity concerns remains strong. A 13% increase in good quality ratings was achieved for the subsection as whole, bringing the total percentage of good quality responses from 33% to 46%. The inclusion of equity has particularly improved since 2010, climbing from 9% in 2010 to 41% in 2012.

Conclusions

Evidence drawn from this and previous review cycles shows an unmistakable trend towards improved report submissions. Over half of reports (62%) meet UNICEF's quality standards (with a rating of highly satisfactory or outstanding), which represents a 20% increase from 2011.

When disaggregating reports by their typology, the extent to which a report satisfactorily met UNICEF standards for high quality is not linked to the nature or focus of the report, but rather on whether it successfully complied with the relevant requirements.

The quality of the reports reviewed did not vary dramatically across sections, although there were some sections that came out stronger than others. Though the methodology section improved when compared to last year, it was noted to be the weakest overall. While ethics was also found to be a particular area of weakness, it too has improved since 2011, with a 10% increase in good quality ratings. The section on the object of the evaluation came out strongest, with 70% of reports adhering to UNICEF's quality standards.

Human rights, gender and equity continue to improve year over year. Although there is relatively little variation between the three areas with regards to their rating levels this year, the integration of equity into evaluation reports has increased dramatically since 2010. This may come as a result of UNICEF's efforts to strengthen internal capacity in equity.

Evaluation reports tend to reflect what evaluators were asked to consider, including the suggested methods, evaluation criteria, data sources, and questions underscored in the ToRs. As such, even though gender, human rights and equity concerns may be central to the values of the Agency, whether or not these are addressed in the evaluation reports tends to be dictated by the ToRs. By the same token, the

extent to which ethics, M&E frameworks, contribution and attribution, evaluation criteria, cost analysis, the presence or absence of conclusions and lessons, and the executive summary or report contents are clearly discussed or presented largely depends on the guidance given to evaluators.

Recommendations

UNICEF should continue to systematically communicate the GEROS results as part of its effort to incentivize managers regarding the system, as well as communicate the specific criteria of GEROS to evaluators.

UNICEF uses several means of communicating the GEROS findings to its staff, including through regional dashboards and the intranet. These efforts should be continued and enhanced, particularly in regions where participation is low. To maximize the effectiveness of the GEROS system, greater efforts should be made to systematically communicate to evaluators the exact quality standards that are used for rating evaluation reports. In this regard, the ToRs could communicate that the evaluation report will be rated by GEROS, and guidance on the GEROS criteria should be provided (either through a website link or a list of the GEROS criteria).

UNICEF’s internal learning systems around evaluation should continue to be strengthened; the GEROS system can play a role in informing the continuous improvement of that learning system.

The UNICEF Evaluation Office has a strong emphasis on learning and has used multiple approaches to foster learning and help build evaluation capacity in regions. Particularly, the evaluation support (such as the Regional Facility) provided by Regional Office M&E advisors to Country Offices seems to be working well, given that such support is developing processes to help improve the pool of qualified evaluators, as well as the quality of ToRs, evaluation workplans and draft reports.

The GEROS meta-evaluation provides insights into certain topics that may require additional attention in such efforts from both HQ and regions: adequately integrating HRBAP, gender equality and equity in the evaluation design, implementation and final reports; ensuring the ToRs reflect the desired criteria and content; focusing on ethics; and encouraging the inclusion of matrices to reinforce linkages between methodology and evaluation questions. Further, appropriate guidance should be given to evaluation managers on how to ensure that their evaluations are properly resourced in terms of the allocation of person days to particular tasks (especially analysis and report drafting, which are often under-resourced). Guidance would include preparing a realistic level of effort according to the existing budget.

UNICEF should continue to review and continually improve the standards used in the GEROS process, even if this risks compromising comparability of GEROS data from year to year.

UNICEF reviews the GEROS criteria (template) on a periodic basis. This is a good practice that should be strengthened so that the review occurs on a more regular basis. In the next review of the GEROS criteria, the Evaluation Office should consider explicitly identifying and clarifying criteria that are ambiguous and/or that may not be applicable for all types of evaluations. It also may wish to review issues that emerge with the change in wording of the rating scale.

Lessons learned

According to UNICEF, lessons learned are contributions to general knowledge that refine or add to common understanding, and should not be merely a repetition of common knowledge.

The general characteristics of a strong evaluation report include clearly and directly addressing the evaluation criteria, good structure, and logical linkages threaded throughout. Thus, while content is important, the presentation of that content is just as important.

The content of an evaluation report is the most intuitively obvious basis for creating a strong report; however, the GEROS process made it clear that *how* information is organized and presented is also a key

factor in the achievement of UNICEF's quality standards. A good report moves between theoretical elements and data that supports those elements, as well as between achievements (outputs) and results (outcomes and impacts), without losing track of the causal variables that affect performance.

Monitoring the quality of evaluations through a GEROS-type system improves the quality of evaluations.

There has been consistent improvement in the quality of UNICEF's evaluations over the last four years, and 2012 was no exception. This underscores that such monitoring of the evaluation report quality is an important element to ensure better evaluation reports.

The more that UNICEF makes clear to evaluators the priorities and foci of its evaluation system, the more likely it is that evaluation reports will meet those standards.

Evaluators usually attempt to satisfy the ToRs and adhere to UNICEF's evaluation standards. If evaluators are aware that they will be judged according to the GEROS standards and know what these standards specify, they will strive to meet these criteria and thereby produce better evaluation reports.

Strong evaluation reports depend upon appropriate time being allocated to analysis and writing.

Based upon the GEROS analysis as well as Universalialia's over 30 years of evaluation experience, strong evaluation reports require time to gather the right data. More importantly, a significant amount of time for data analysis and report writing is also needed. Allocating appropriate time for these latter stages, based on the evaluation's size, complexity and scope, will tend to produce better reports.

Acronyms

CEE/CIS	Central and Eastern Europe/Commonwealth of Independent States (Regional Office)
CCC	UNICEF's Core Commitments for Children
CEDAW	Committee on the Elimination of Discrimination against Women
COs	Country Offices
CRC	Convention on the Rights of the Child
EO	Evaluation Office
EAPRO	East Asia and Pacific Regional Office
EQOS	Evaluation Quality Oversight System
ESARO	East and Southern Africa Regional Office
GEROS	Global Evaluation Report Oversight System
GE	Gender Equality
HRBAP	Human Rights Based Approach to Programming
HQ	Headquarters
M&E	Monitoring and Evaluation
MENARO	Middle East and North Africa (Regional Office)
MfDR	Managing for Development Results
MTSP	Medium Term Strategic Plan
NA	Not Applicable
RBM	Results Based Management
ROs	Regional Offices
ROSA	Regional Office of South Asia
RTE	Real-time evaluation
ToRs	Terms of Reference
TACRO	The Americas and Caribbean Regional Office
UNDAF	United Nations Development Assistance Framework
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
WCARO	West and Central Africa Regional Office
CEE/CIS	Central and Eastern Europe/Commonwealth of Independent States (Regional Office)
CCC	UNICEF's Core Commitments for Children
CEDAW	Committee on the Elimination of Discrimination against Women

Contents

1. Introduction	1
2. Background to Geros	1
3. Purpose, Objectives & Scope of Geros	2
4. Methodology	2
4.1 Review of Evaluation Reports	2
4.2 Meta-evaluation	5
4.3 Changes Made From Previous Years	5
4.4 Limitations	6
5. Findings	6
5.1 Overall Ratings and Feedback	6
5.2 Overall Regional Trends	9
5.3 Trends by Type and Scope of Evaluation	10
5.3.1 Geography	10
5.3.2 Management of Evaluation	11
5.3.3 Purpose of the Evaluation	12
5.3.4 Results	13
5.3.5 Stage	15
5.3.6 MTSP Correspondence	16
5.3.7 Level of Independence	17
5.3.8 Trends by Language of Report	17
5.3.9 Coherence	18
5.4 Trends by Quality Assessment Category	18
5.4.1 Overall trends Sections A-F	18
5.4.2 Section A: Object of the Evaluation	19
5.4.3 Section B: Evaluation Purpose, Objectives and Scope	21
5.4.4 Section C: Evaluation Methodology, Gender, Human Rights and Equity	23
5.4.5 Section D: Findings and Conclusions	26
5.4.6 Section E: Recommendations and Lessons Learned	28
5.4.7 Section F: Report structure, logic and clarity	30
5.5 Examples of Good Practices	31
6. Conclusions	32
7. Recommendations	33
8. Lessons Learned	34

Exhibits

Exhibit 4.1	Performance Scorecard	3
Exhibit 4.2	Changes in Coding	5
Exhibit 5.1	Reports Reviewed per region per year	7
Exhibit 5.2	Good Quality Evaluation Reports Over Time (2009-2012)	7
Exhibit 5.3	Overall Ratings for 2012	8
Exhibit 5.4	Inclusion of ToRs in Reports Over Time (2009-2010)	8
Exhibit 5.5	Proportion of Good Quality Reports Per Region	9
Exhibit 5.6	Proportion of Reports Reviewed and Good Quality Reports By Geographic Scope (2012)	10
Exhibit 5.7	Report Quality Progression 2010-2012 (% of good quality reports)	11
Exhibit 5.8	Proportion of Reports Reviewed and Good Quality Reports by Management of the Evaluation (2012)	11
Exhibit 5.9	Good Quality UNICEF-Managed Evaluations over Time	12
Exhibit 5.10	Proportion of Reports Reviewed and Good Quality Reports by Purpose (2012)	13
Exhibit 5.11	Proportion of Reports by Results Level Over Time	14
Exhibit 5.12	Proportion of Reports Reviewed and Good Quality Reports by Results Level (2012)	14
Exhibit 5.13	Proportion of Reports Reviewed and Good Quality Reports by Stage	15
Exhibit 5.14	Proportion of Reports Reviewed and Good Quality Reports by MTSP Correspondence	16
Exhibit 5.15	Proportion of Reports by Level of Independence (2009-2012)	17
Exhibit 5.16	Proportion of Reports Reviewed and Good Quality Reports by Language	18
Exhibit 5.17	Good Quality Ratings per Section- Year by Year Progression (2010-2012)	19
Exhibit 5.18	Section A Ratings and Comparison with 2011	20
Exhibit 5.19	Sub-Section Ratings: Object of the Evaluation	20
Exhibit 5.20	Spread of results of cornerstone questions	21
Exhibit 5.21	Section B Ratings and Comparison with 2011	22
Exhibit 5.22	Sub-Section Ratings: Purpose, objectives and scope of the evaluation	22
Exhibit 5.23	Section C Ratings and Comparison with 2011	23
Exhibit 5.24	Sub-section ratings: Methodology, Gender, Human Rights, and Equity	24
Exhibit 5.25	Inclusion of Human Rights, Gender, and Equity: Good Quality Ratings Year by Year	26
Exhibit 5.26	Section D Ratings and Comparison with 2011	26
Exhibit 5.27	Sub-section ratings: Findings and Conclusions	27
Exhibit 5.28	Section E Ratings	28
Exhibit 5.29	Sub-Section Ratings: Recommendations and Lessons Learned	29
Exhibit 5.30	Section F Ratings and Comparison with 2011	30
Exhibit 5.31	Sub-Section Ratings: Report Structure, Logic and Clarity	30

Appendices

Appendix I List of Findings	37
Appendix II List of Recommendations	39
Appendix III Terms of References	41
Appendix IV List of Reports Assessed	45
Appendix V Review tool	51
Appendix VI Regional Ratings Graphs	69
Appendix VII UNICEF, Country-led and Jointly Managed Evaluations by Region	73
Appendix VIII Overall Ratings Graphs by Report Typology	77
Appendix IX Sub-section ratings	81
Appendix X Analysis Table for 2012 Reports	85
Appendix XI Clarification of criteria for completing review	89

1. Introduction

The Universalialia Management Group Limited (Universalialia) is pleased to submit to UNICEF the final Global Meta-Evaluation Report on 2012 evaluation reports. As part of the overall Global Evaluation Reports Oversight System (GEROS) project, the purpose of this report is to summarise key trends, weaknesses and strengths, as well as draw lessons learned and good practices emerging from our review of 79 reports carried out in 2012. In an effort to support continuous learning within UNICEF, this report also includes an analysis of the GEROS process and provides suggestions for future improvements.

The report contains both quantitative analyses of the number of items reviewed and the average ratings, as well as qualitative assessments on observed strengths and areas for improvements. Throughout, examples of good practices are highlighted and where applicable, suggestions for ways to further improve these are presented. The report ends with a review of the major conclusions of the meta-evaluation, lessons learned and recommendations.

This report is organised as follows:

- Section 2 – Background to GEROS
- Section 3 – Purpose, Objectives & Scope of GEROS
- Section 4 – Methodology
- Section 5 – Findings
- Section 6 – Conclusions
- Section 7 – Recommendations
- Section 8 – Lessons Learned

The Appendices provide supplementary material, including the Terms of Reference and the GEROS assessment tool.

2. Background to GEROS

UNICEF's establishment of a Global Evaluation Quality Assurance System to ensure that evaluations managed or commissioned by UNICEF meet high quality standards has its roots in the Global Evaluation Compact endorsed in 2009, which commits the Evaluation Office (EO) and the Regional Offices (ROs) to collaborate in strengthening the evaluation function within UNICEF. The Evaluation Policy and Executive Directive set out clear roles and responsibilities as well as management measures to strengthen the evaluation function in the organisation.

Because of the decentralized nature of UNICEF, the majority of evaluations supported by the agency are managed at the Country Office level. While this ensures that evidence generated is relevant to local contexts – and therefore more likely to inform national policies for children – such decentralization presents challenges in terms of establishing a system that can ensure good quality, high credibility, and utility that are consistent across the organisation. To address such variations across regions and Country Offices in the way M&E functions are carried out, the Global Evaluation Report Oversight System (GEROS) was developed to provide greater oversight of the Evaluation function within UNICEF. Following three years of implementation, GEROS underwent another review in 2012 that involved the Regional Offices and the EO.

Capacity building efforts at the regional and Country Office levels remain ongoing; for example, some Regional Offices (such as CEE/CIS) have established a regional facility to support the review of ToRs, draft inception reports and draft evaluation reports. Such efforts, in concert with insights from past meta-evaluations, are helping UNICEF to better monitor progress, highlight emerging strengths, and identify areas for further improvement.

3. Purpose, Objectives & Scope of GEROS

GEROS has four main objectives:

- To provide senior managers in the UNICEF HQ, Regional Offices, and Country Offices with a clear and succinct independent assessment of individual evaluation reports;
- To strengthen internal evaluation capacity by providing feedback to commissioning offices with practical recommendations to improve future evaluations and to inform their own assessment of the performance of external consultants who might be hired for future evaluations;
- To report on the quality of evaluation reports through a review and assessment process whose results are communicated to senior management and inclusion of this information in the Global Evaluation Database; and
- To contribute to the EO's corporate knowledge management and organisational learning, by providing the evidence base for a meta-analysis of good quality reports.

These objectives are embedded in the conception of GEROS as an organisation-wide system that is complemented by other quality-assurance mechanisms upstream in the evaluation process, such as the use of Terms of Reference (ToRs) and Report Checklists. It is grounded in the standards developed by the United Nations Evaluation Group (UNEG) and incorporates UNICEF-specific standards related to the agency's commitment to integrate gender equality considerations, a focus on equity, and human rights-based approaches, particularly child rights.

Carried out by an independent panel of experts to ensure the credibility of the review process, GEROS provides a rigorous framework for carefully and methodically assessing the quality of the reports commissioned by the various units of the agency, including country, regional and headquarters. A copy of the template used as part of the GEROS review process is provided in Appendix V of this report.

4. Methodology

The evidence discussed in this report is derived from three distinct project phases. The first phase involved a two-day training event with the review team (with the collaboration and participation of UNICEF), in order to provide the team with hands-on practice and clarification for each of the sections of the evaluation tool. The second phase involved a systematic review of all UNICEF sponsored evaluations completed in 2012 using the GEROS template to assess the quality of submitted reports. Finally, a meta-evaluation of all submitted reports was completed. The review methodology used in the second phase of the assignment is presented in section 4.1 below and the methodological approach taken for the meta-evaluation itself is presented in section 4.2.

4.1 Review of Evaluation Reports

UNICEF evaluation reports were assessed using a two-pronged approach, involving: (i) the systematic application of the GEROS evaluation tool to each report, and (ii) the use of a quality assurance process.

Review Process and Evaluation Tool

Reports submitted to the UNICEF Global Evaluation Database for 2012 were shared with the review team over a four-month period. Submitted reports were first screened by the EO to ascertain that they were

evaluations² before being screened again by the review team prior to commencing the analysis. This secondary screening process was used to classify evaluation reports in terms of their geographical coverage, managerial oversight, purpose, levels of change sought, the Medium Term Strategic Plan MTSP correspondence, level of independence, and stage of evaluation (i.e. summative, formative or both). This additional screening also helped to identify reports that were not in fact evaluations. As a result of these screening processes, a total of 79 reports were reviewed by Universalia between mid-January and the end of May 2013.

Upon reception of the reports, each evaluation was assigned to a reviewer on the basis of thematic expertise and linguistic abilities. Reviewers assessed the merits of the reports based upon the UNICEF-adapted UNEG evaluation reports standards³ set forth in an excel-format evaluation tool. In all, 58 guiding questions and two interdependent tiers of aggregation were used to filter reports. At the macro level, the tool had six sections, covering:

- (i) The object of the evaluation;
- (ii) The evaluation purpose, objectives and scope;
- (iii) The evaluation methodology, gender, and human rights;
- (iv) Findings and conclusions;
- (v) Recommendations and lessons learned; and
- (vi) The extent to which the report was well structured, logical, and clear.

These sections are then divided into 22 sub-sections associated with key performance areas such as the theory of change, the evaluation framework, and methodological robustness. The evaluation team assigned each question a grade using a colour coded four point scorecard (see Exhibit 4.1 below), and provided a justification for the assigned grade in a narrative analysis.

Exhibit 4.1 Performance Scorecard

Colour coding	CC	Dark green	Green	Amber	Red	White
	Questions		Outstanding	Yes	Mostly Satisfactory	No
Section & Overall Rating		Outstanding, best practice	Highly Satisfactory	Mostly Satisfactory	Unsatisfactory	

The overall approach to report analysis was qualitative in nature. This qualitative approach was designed to enable reviewers to provide useful analysis across the range of evaluation contexts encountered, as well as constructive feedback to improve future evaluation reports. As shown in Exhibit 4.1 above, each section and the overall report were given a rating of either: “Outstanding, Best Practice”, “Highly Satisfactory”, “Mostly Satisfactory”, or “Unsatisfactory”, while each individual question was rated either

² The Global Evaluation Database contains not only evaluations, but also surveys, studies and various other research contributions. However, as defined by UNICEF itself, an evaluation incorporates “*judgment [on] the relevance, appropriateness, effectiveness, efficiency, impact and sustainability of development efforts, based on agreed criteria and benchmarks among key partners and stakeholders. It involves a rigorous, systematic and objective process in the design, analysis and interpretation of information to answer specific questions. It provides assessments of what works and why, highlights intended and unintended results, and provides strategic lessons to guide decision-makers and inform stakeholders.*” (UNICEF, 2011)

³ The eight UNICEF-adapted UNEG Evaluation reports standards are closely linked to the sections of the Geros template. They include: (i) the report structure; (ii) the object of evaluation; (iii) the evaluation purpose, objective(s) and scope; (iv) the evaluation methodology; (v) findings; (vi) conclusions and lessons learned; (vii) recommendations; and (viii) gender and human rights, including child rights (see Annex 2 of the Global Evaluation Reports Oversight System, December 2012).

“Outstanding”, “Yes”, “Mostly” and “No”. An “N/A” option was also available, when a particular question was not relevant because of the nature of the report.

In addition to ratings, a commentary was provided for each section and sub-section, suggestions for future improvements were provided for each section, and executive feedback was provided for each section and for the overall report. The complete review process generated three types of data: a report typology, a series of ratings, and a structured set of comments on the quality of the report, based on corresponding questions, sub-sections, and sections.

Each report is thus associated with the described performance scorecard that was used to compare results across entire populations of reports reviewed. Senior evaluation team members provided on-going support to the review team and reviewed more than half of their assessments to ensure completeness, clarity and consistency of analyses, as well as make sure that executive summaries and feedback were aligned with UNICEF’s priorities. The overall rating of each completed review was entered into a database comprising all of the evaluations included in the review process. This database provided UNICEF with periodic updates on progress made and gave a clear indication of reviews completed and submitted to the organization.

Quality Assurance Process

The Geros process was carried out by a dedicated project team composed of reviewers, senior reviewers for quality assurance, and a project manager. Senior members were tasked with overseeing, coordinating, and supporting the review process. Through this structure, the team of reviewers received continuous support from senior team members, who in turn helped to ensure the rigour and validity of the reviews. As a quality assurance method, overlapping levels of oversight helped to combine evaluation expertise, language skills, sectoral expertise, database management experience, and knowledge of UNICEF and of UNICEF’s context, including its programmatic and strategic objectives. The reliability of the review process was assured through the following steps:

- **Training:** Members of the review team were trained prior to commencing the reviews in how to use assessment tools and templates. The purpose of this step was to ensure that all team members developed a common understanding of the process and standards of delivery.
- **Senior and Peer Reviews:** Prior to submitting completed assessments to UNICEF, report reviews were submitted to a senior team member to ensure completeness, quality, and adherence to established standards. Though Universalialia had anticipated that only 20% of the reports would be reviewed by senior project team members (as suggested in the Geros guidance document), our early experience showed that confidence in the completed analyses could best be achieved through the corroboration of a second reading.

So while senior team members tended to concentrate their attention on the reports that were flagged by UNICEF for their strategic significance as well as on those reports that received lower levels of quality, Universalialia in parallel initiated a peer review system to validate initial assessments, ensure consistency in the application of assessment standards (i.e. ensuring reliability and consistency between raters), flag discrepancies or areas in need of further clarification, and make editing corrections as required. Consequently, every report was reviewed at least twice.

- **UNICEF Feedback:** Consistency in the application of ratings, the level of analysis provided and the overall results of each report were closely monitored by the UNICEF Evaluation Office. As a result of this process, a handful of reports were sent back to the review team for further analysis and/or more detailed explanations on the factors that supported a particular rating.

4.2 Meta-evaluation

The meta-analysis presented in this report was derived from a total population of 79 reports, covering all of the evaluation reports from 2012 submitted to the UNICEF database. Findings for this report were generated as follows:

First, once all reviews had received the final quality check from UNICEF, Universalialia built a Visual Basic for Applications (VBA) code in Excel to aggregate the ratings and comments of each review file into a single excel file, thus generating a set of data to be used in the analysis of overall trends, strengths and weaknesses of all of the evaluation reports reviewed.

Second, in order to facilitate comparison with previous reports, trends were analysed the same way as in previous years. Thus, as in last year’s meta-evaluation, overall ratings as well as each individual section’s ratings were disaggregated according to the report typology (region, geographic scope, management, purpose, result-level, level of independence, stage, medium-term strategic plan focus areas, and language of the report). As a result, trends were analysed by report typology and each section was catalogued on the basis of the four-point scale detailed in Exhibit 4.1, where outstanding and highly satisfactory ratings were often grouped together to show the totality of good quality reports that adhere to UNICEF standards. Tabulated results were converted to percentage ratios, and data from 2009-2011 was used to create year to year comparisons and thus report on improvements overtime, as well as areas in need of improvement. A series of graphs were created to visualize the data and include it in the report. The table used to analyse the trends can be found in Appendix X.

Third, a combination of qualitative and quantitative data was used in order to analyse the trends in each individual section rating (i.e. context and object of the evaluation, methodology, etc). In order to determine the major strengths and weaknesses and to foster a deeper understanding of these, the ratings for each individual question were disaggregated by their corresponding sub-section (i.e. theory of change, evaluation framework, data collection), and the proportion of the ratings per sub-section were calculated in order to analyse the overall areas of strength and weakness. In addition, reviewer comments were compiled and disaggregated through key word and corollary term searches to identify emerging themes that complement the quantitative data.

4.3 Changes Made From Previous Years

The Geros process was initiated in 2010 (covering reports written in 2009) and though standards and evaluative questions remain largely unchanged, some aspects of Geros were changed in 2012 after the rapid review carried out by UNICEF.

- The four point scale used to grade reports was changed from a “confidence scale” to a rating system that retained the four point scale but used the level of satisfaction to convey overall levels of appreciation (see Exhibit 4.2 below). For consistency purposes, the color coding has remained unchanged.

Exhibit 4.2 Changes in Coding

Dates	Dark Green	Green	Amber	Red	White
2009-11	Very confident to act	Confident to Act	Almost confident to act	Not confident to act	Not Applicable
2012-15	Outstanding	Highly Satisfactory	Mostly Satisfactory	Unsatisfactory	

- Two new options were added to the classification of “Management of the evaluation”:
 - “Externally managed: An external organization manages the evaluation, where UNICEF is amongst the organizations being assessed (UN and non-UN)”. This option was added as a result of several evaluation reports included in the Geros internal system that were not managed by UNICEF, but rather by other organizations.
 - “UNDAF Evaluation: They are a joint UN review, conducted with national partners, about the overall results expected from the UN cooperation in the country”. This new option was added to differentiate these evaluations from any other types of jointly managed evaluations.
- Another option was added to the classification of “Purpose” of the evaluation, which refers to “regional/multi-country programme evaluation” and includes evaluations that assess several programmes from a regional or multi-country perspective.
- A new option was added to the classification of the evaluation’s stage, “summative and formative” to reflect the fact that many evaluations are done for summative as well as formative purposes.

4.4 Limitations

The evaluation team faced a number of issues in conducting this meta-evaluation. First, as noted in previous meta-evaluations, the Geros process relied on the subjective application of objective criteria to make judgments on quality or worth of an evaluative contribution. Reviewers were tasked with the difficult process of making judgments based solely on the reports themselves since they did not have access to the evaluator’s data or the programme officer’s knowledge of the issues explored, and only limited access to additional sources of information to validate the inferences made in the report. Moreover they must do so on the basis of the Geros template and not on the ToRs. These issues may have affected some evaluations negatively if the requirements of the Geros template were satisfied in for example the inception report but not in the final report.

The other main limitation related to the interpretation of key questions and issues within the Geros framework. For instance, issues such as the level of independence of the evaluation, inclusion of HRBAP, or the extent to which an attempt was made to construct a counterfactual prompted discussion within the review team. While the review team worked assiduously to ensure consistency of ratings through a rigorous overlapping checking process and particularly through the peer reviews, it is possible that there were some inconsistencies on particular questions.

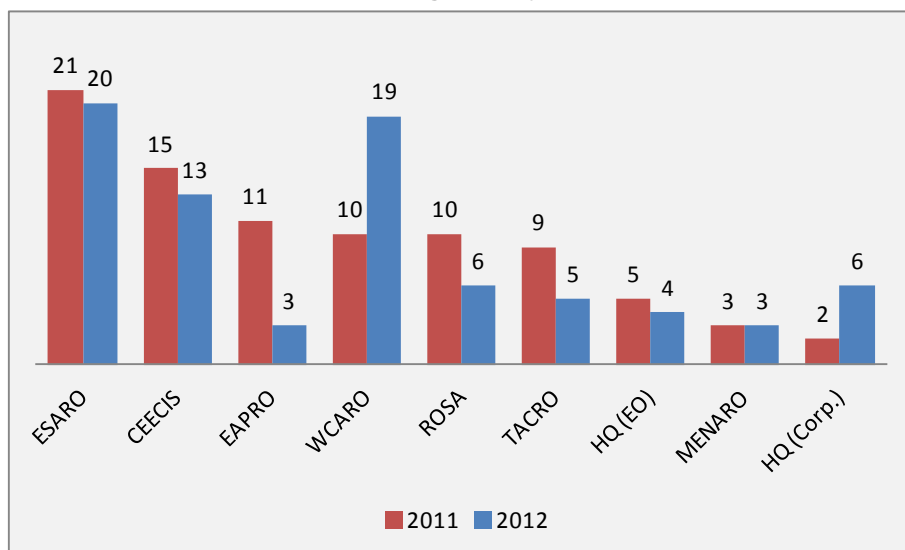
Finally, it should be noted that the data for 2009 was limited, and thus a comparison across all four years of implementation was not always possible. Moreover, as noted earlier (see Section 4.3), the scale used to measure achievement of standards was changed for the year 2012; while the spread of the four-point scale did not change, the language did. Though in theory this should not affect rating comparability since it maintains the spread of probable answers along the same four point scale, there may be a different attachment to the meaning of the new scale’s language.

5. Findings

5.1 Overall Ratings and Feedback

As shown in Exhibit 5.1 below, the regional distribution pattern of completed reports is not always the same as in 2011. Although East and Southern Africa Regional Office (ESARO) and the Central and Eastern Europe/Commonwealth of Independent States Regional Office (CEECIS) continue to lead in terms of number of reports submitted and reviewed, the West and Central Africa Regional Office (WCARO) significantly increased the number of evaluations submitted to the database and reviewed.

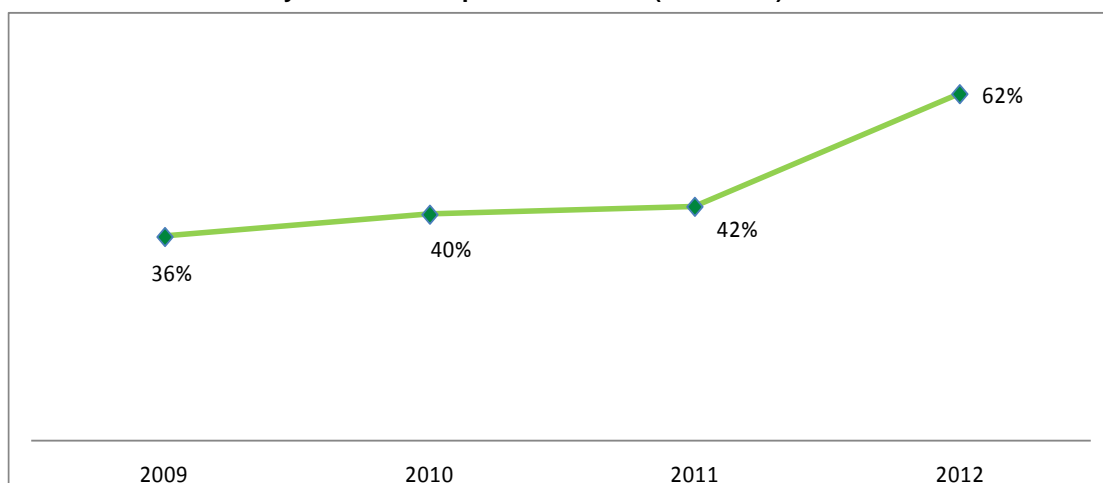
Exhibit 5.1 Reports Reviewed per region per year



Finding 1: The quality of reports submitted to UNICEF increased sharply between 2011 and 2012.

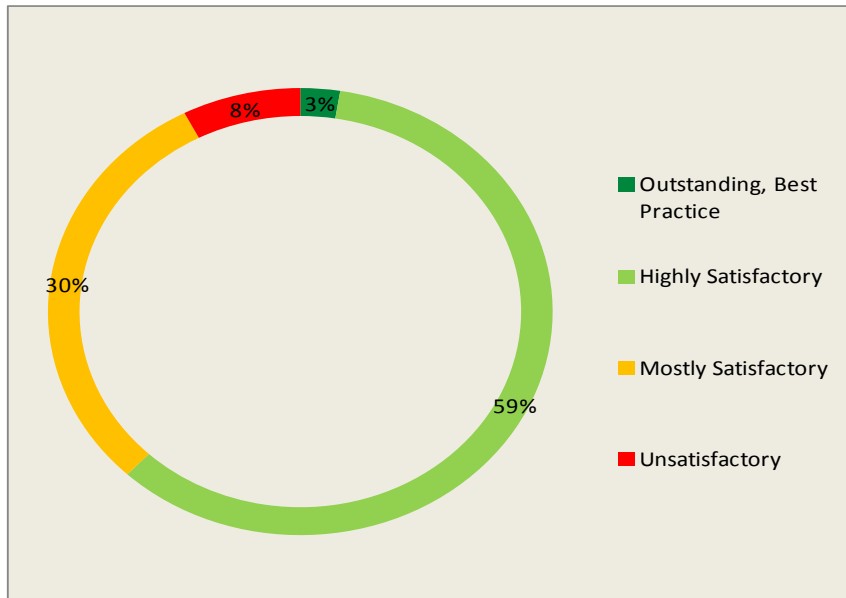
Overall results of the meta-analysis point to a net improvement in the quality of the reports submitted to UNICEF. As shown in Exhibit 5.2 below, the percentage of good quality reports (those rated as highly satisfactory or outstanding) rose by 20 percentage points between 2011 and 2012, from a previous high of 42% to 62%. This is a significant and positive change, though several more iterations of the yearly analyses will be required before generalised improvements in the quality of submitted reports can be definitively confirmed.

Exhibit 5.2 Good Quality Evaluation Reports Over Time (2009-2012)



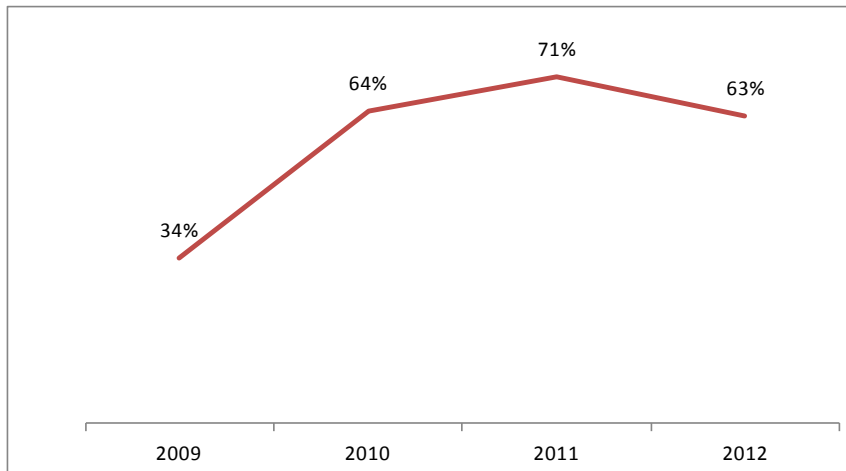
The overall ratings for 2012 are shown in Exhibit 5.3 below. In 2012, only 8% of reports were rated as unsatisfactory (compared with 23% in 2011); 30% were found to be mostly satisfactory (instead of 35% in 2011), and 59% were considered highly satisfactory (compared to 37% in 2011).

Exhibit 5.3 Overall Ratings for 2012



The ratings of mostly satisfactory and unsatisfactory (yellow and red categories) ratings often reflect that the evaluation report had not addressed a large proportion of the GEROS criteria and questions. More often than not, the problem was that ToRs did not require all of these elements. However, ToRs are often not included as part of the evaluation report (see Exhibit 5.4) so it is difficult to cross-check the expectations as expressed in the ToRs with the issues actually addressed in the evaluation report. Although no direct correlation can be established between the inclusion of the ToRs and the quality of the report, review data shows that reports with annexed ToRs tended to be rated better (70% were rated as highly satisfactory or outstanding compared to only 48% that received these ratings for reports that did not include them).

Exhibit 5.4 Inclusion of ToRs in Reports Over Time (2009-2010)⁴



⁴ The Inclusion of ToRs by region for 2012 is in Annex VI.

5.2 Overall Regional Trends

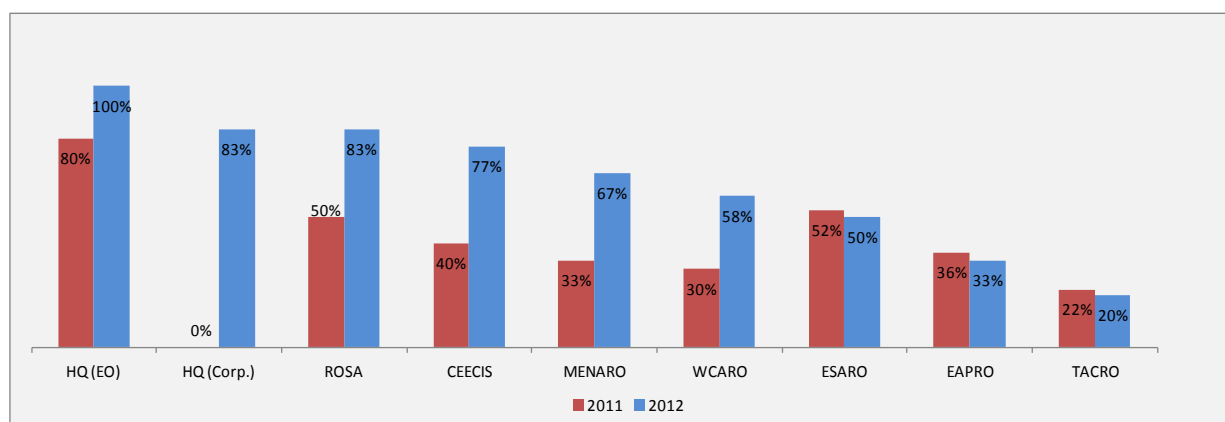
Finding 2: With few exceptions, the quality of UNICEF sponsored evaluations has improved across most of the Regional Offices when compared with previous years.

When results from 2012 are compared to analyses from previous years, regional trends point to a steady progression in the quality of submitted reports.

In considering the proportion of reports that successfully met UNICEF standards, notable improvements were seen across the majority of Regional Offices, and the proportion of reports rated as unsatisfactory diminished across all regions. In some cases, the registered gains were substantial. As shown in Exhibit 5.5, the proportion of favourable reports nearly doubled for the Regional Offices in Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS), the Middle East and North Africa (MENARO), and West and Central Africa (WCARO), increased by 33% and 20% for the Regional Office for South Asia (ROSA) and HQ (EO) respectively, and achieved an 83% success rate in the case of HQ (Corp.), which had not produced a single outstanding or highly satisfactory report the year before. When taken together, HQ (EO-led) and HQ (Corporate), the increase in the level of quality is also significant, with 90% rated as good quality, compared to 57% last year. Another revealing element of the positive trends uncovered in the current cycle of reviews pertains to the overall proportion of reports that were rated as mostly satisfactory and/or unsatisfactory. In considering the regions that produced the greatest number of reports, we note that in the case of WCARO for instance, the proportion of reports that achieved lower levels of confidence (70%) in 2011 had fallen by 28 percentage points to just 42%. Similarly, the proportion of weaker reports in CEE/CIS dropped from a high of 60% in 2011 to only 23% in 2012. While the overall ratio of poorer performing reports for ESARO remained nearly identical between 2011 (48%) and 2012 (50%), the number of reports rated as unsatisfactory dropped significantly (from 24% in 2011 to 5% in 2012). As in 2011, the two regions with the lowest performance levels were the Regional Offices of East Asia & Pacific (EAPRO) and The Americas and Caribbean (TACRO), which respectively achieved only 34% and 20% rates of overall good quality reports. It is worth noting however, that caution must be taken when interpreting results for the regions that did not conduct many evaluations, such as MENARO and EAPRO, where only 3 reports were reviewed this year.

Taken together, these results suggest that Geros is having a positive effect on the overall completeness and conformity of evaluation reports year by year. The evidence also points to particular regions where additional assistance may be warranted to ensure further improvements in the quality of UNICEF sponsored evaluations.

Exhibit 5.5 Proportion of Good Quality Reports Per Region⁵



⁵ The percentage of good quality reports was calculated by adding the number of reports that were outstanding and highly satisfactory per region over the total number of reports per region. Total number of reports per region can be seen in Exhibit 5.1

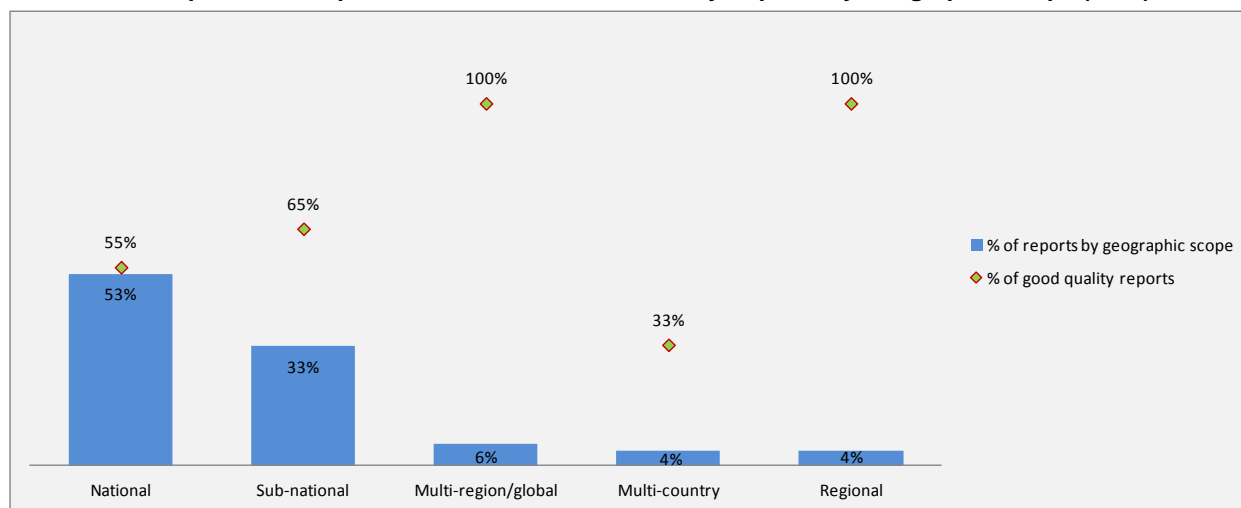
5.3 Trends by Type and Scope of Evaluation

5.3.1 Geography

Finding 3: The geographic scope of most evaluations continues to be at the national or sub-national level. However, the percentage of good quality reports has proportionally increased for all of the categories of geographic scope.

Classification by geography refers to the coverage of the programme being evaluated and generalizability of evaluation findings. As in previous years, the proportion of reports with a national and sub-national scope constituted the vast majority of reports submitted (53% national and 33% sub-national for 2012 – See Exhibit 5.6). But unlike in previous years, the percentage of sub-national reports rated highly satisfactory or above nearly doubled from 33% in 2011 to 65% in 2012. There were also notable improvements in the ratio of reports that achieved good quality ratings at the national (from 47% in 2011 to 55% in 2012), regional (increase from 75% to 100%), and multi-regional or global levels (increase from 67% to 100%). The only noted decline in overall levels of quality was found in relation to reports that evaluated multi-country programs, which dropped from a 50% to 33% good quality rating between 2011 and 2012. Yet with so few reports for this multi-country category (two in 2011 and three in 2012), as well as for the regional and multi-region/global category, a single review can have a disproportionate impact on the category’s performance.

Exhibit 5.6 Proportion of Reports Reviewed and Good Quality Reports By Geographic Scope (2012)⁶



The trends for national and sub-national reports suggest a continuous progression. As shown in Exhibit 5.7, there has been a progressive improvement in the quality of reports. Since these two categories represent a large majority of the evaluations reviewed (86%), it reflects the overall improvement in report quality since 2010.⁷

⁶ Total number of reports per geographic scope: 26 for sub-national, 42 national, 3 multi-country, 3 regional, and 5 multi-region/global.

⁷ The small numbers of reports in the other categories (multi-country, multi-region, regional) limits the possibility of similar analysis.

Exhibit 5.7 Report Quality Progression 2010-2012 (% of good quality reports)

Scope	2010	2011	2012
Sub-National	31%	33%	65%
National	46%	47%	55%

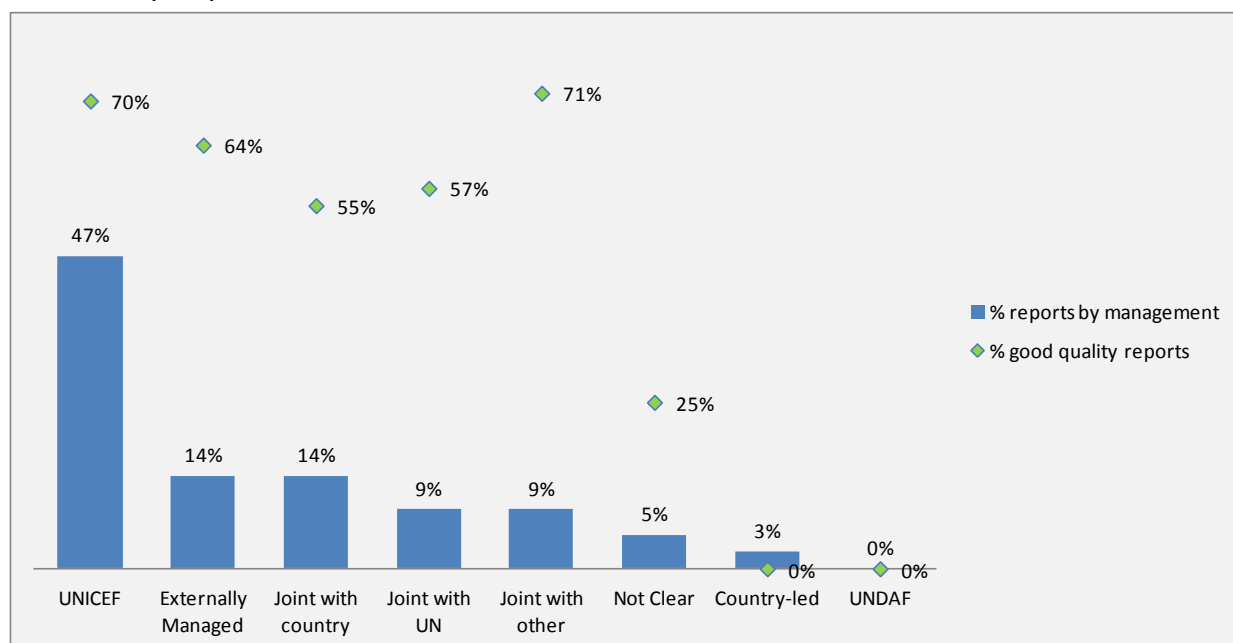
5.3.2 Management of Evaluation

Finding 4: The quality of UNICEF managed or co-managed evaluations has increased markedly since 2011.

While the number of reports for UNICEF-managed evaluations decreased slightly from last year, from 61% of all reports reviewed in 2011 to 47% in 2012, the quality of these reports increased noticeably from 47% with highly satisfactory ratings to 70% (See Exhibit 5.9). Moreover, these positive trends are not exclusive to UNICEF-managed evaluations but were also found across nearly all joint management structures that involved UNICEF. This suggests that UNICEF is becoming increasingly effective at directing evaluative inquiries in ways that support both the interests of the Agency and of its partners (see Exhibit 5.8). Exhibit 5.8 below also indicates that reports that are ambiguous relative to the underlying unit of management also tend to achieve lower levels of quality overall.

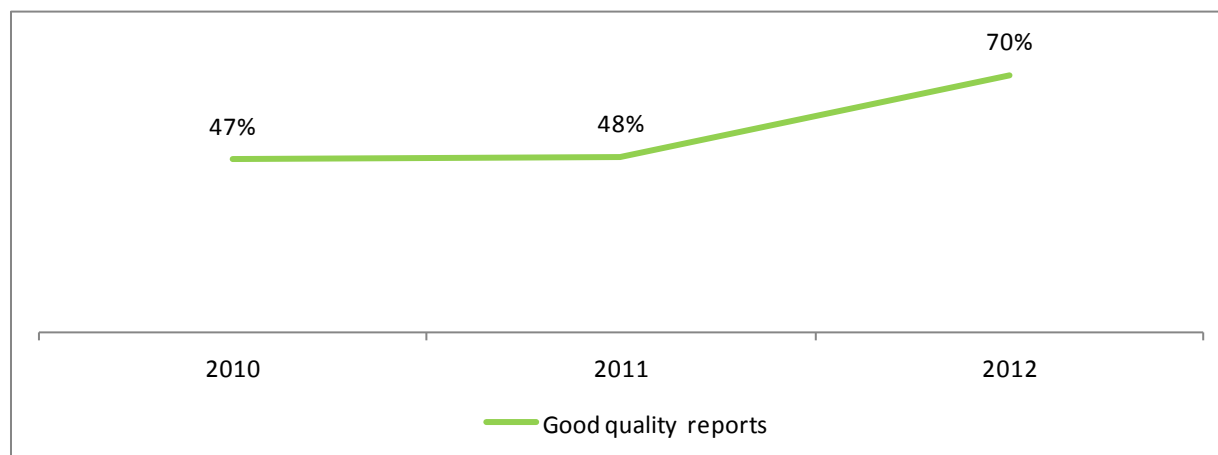
Given the recent importance of country-led evaluations to UNICEF, it is important to note that the number of country-led evaluations continues to be low for 2012 with two reports in total, compared to one submitted in 2011.

Exhibit 5.8 Proportion of Reports Reviewed and Good Quality Reports by Management of the Evaluation (2012)^{8 9}



⁸ The variable “% of satisfactory reports” does not add up to a total of 100%. This is because the variable represents the number of reports rated highly satisfactory and outstanding out of the total number reports per type (e.g. UNICEF, externally managed, etc).

⁹ Total number of reports per management type: 37 for UNICEF, 7 joint with UN, 7 joint with other, 11 joint with country, 2 country-led, 11 externally managed, 0 UNDAF, 4 not clear.

Exhibit 5.9 Good Quality UNICEF-Managed Evaluations over Time


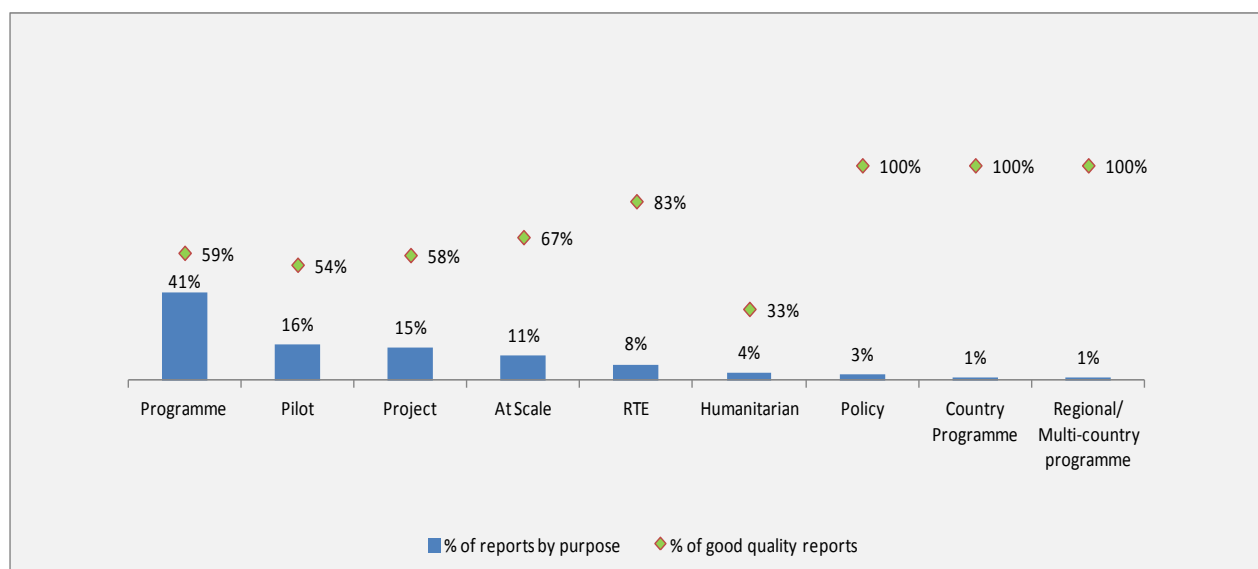
5.3.3 Purpose of the Evaluation

Finding 5: While number of evaluations by purpose – programme, policy, humanitarian, country evaluations, etc. – appear to differ little from year to year, the quality of reviewed submissions has improved for nearly all types of evaluation reports.

The range of purposes addressed in 2012 differed only slightly from previous years (see Exhibit 5.10). Programme evaluation continues to be the most common type of evaluation conducted, increasing from 36% in 2011 to nearly 41% in 2012, along with a one percentage point increase in the levels of conformity with UNICEF standards to 59%. The percentage of evaluated pilot initiatives also increased, from 12% to 16% between 2011 and 2012, as did their overall performance in meeting UNICEF standards, from 40% to 54%. Finally, the number of real time evaluations (RTE) in 2012 increased in number from 1 to 6 (representing a change of 7%), with an 83% compliance rate. While some forms of evaluation were too few in number to draw any useful conclusion (e.g. country, humanitarian or policy evaluations), others such as project and “at scale” evaluations diminished in terms of relative percentages (from 27% to 15% and 16% to 15% respectively) but improved relative to applicable UNICEF evaluation standards (from 30% to 53% compliance for project evaluations and 36% to 67% for the evaluation of at scale initiatives).

Even though no clear relationship can be established between the quality of completed reports and the purpose of the evaluation itself, some trends within the different types of evaluations and reasons for their lower quality can be noted. Looking at the weaker reports, pilot evaluations tended to be especially weak in their evaluation framework, the methodology, and the section on findings. Programme evaluations tended to be weaker when developing recommendations, with 63% of the evaluations not meeting the UNICEF quality standards in this section, and project evaluations tended to have issues with the methodology, with 58% not meeting quality standards for this section.

Exhibit 5.10 Proportion of Reports Reviewed and Good Quality Reports by Purpose (2012)¹⁰



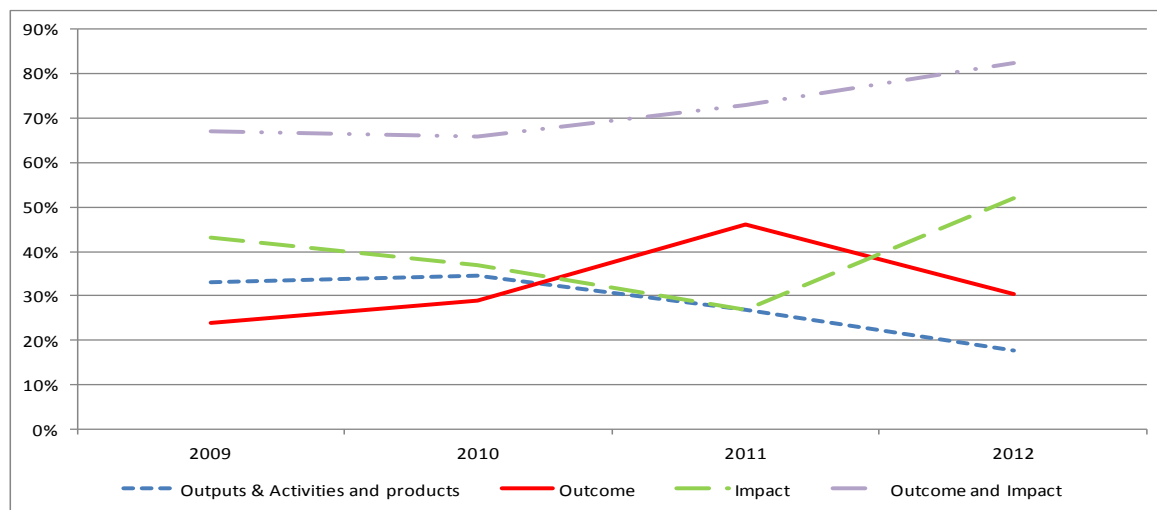
5.3.4 Results

Finding 6: The proportion of evaluation reports considering results at outcome and impact level has fluctuated over time. Reports that focus on output-level results, which are fewer in number and are often humanitarian or real-time evaluations, demonstrate the greatest improvement in terms of quality. Reports that address outcomes continue to underperform with regard to meeting UNICEF quality standards.

The proportion of reports dealing with outcome related results decreased quite significantly between 2011 and 2012 (see Exhibit 5.11), while the proportion of impact level reports increased. The net effect of the combined increase in outcome and impact level assessments can readily be observed in the reduced number of reports that attended to output and activities and products level results (including reports focused on products or activities, such as humanitarian issues or RTEs): from a high of 34% in 2010, the percentage of reports focused on output and activities and products level results dropped by nearly half to 18% in 2012.

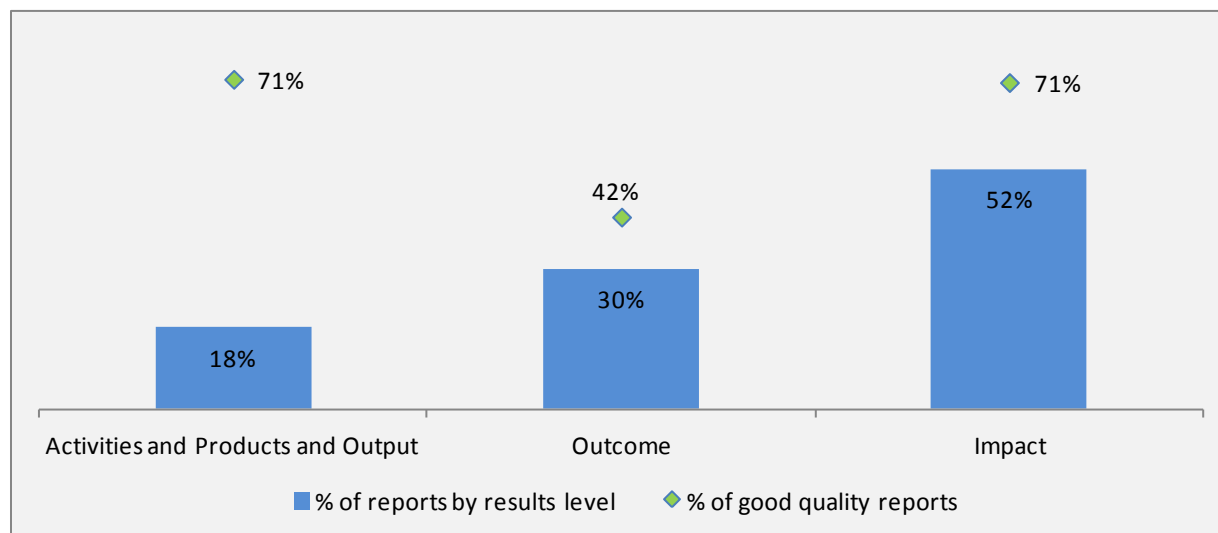
¹⁰ Total number of reports per purpose: 13 pilot, 9 at scale, 2 policy, 6 RTE, 12 project, 32 programme, 1 country programme, 0 joint with country, 1 regional/multi-country programme.

Exhibit 5.11 Proportion of Reports by Results Level Over Time



The proportion of good quality impact level reports increased slightly between 2011 and 2012, from 61% to 71% (see Exhibit 5.12). Outcome-level evaluations received the worst rating, with only 42% of the reports meeting UNICEF standards—largely consistent with last year’s score of 41%. Assessments of outputs, including activities and products, were largely positive, with a combined ratio of 71% of reports achieving outstanding and highly satisfactory ratings (representing a significant increase from 22% in 2011). However, it is important to note that most of the evaluations only reporting lower-level (output) results were humanitarian or real-time evaluations and thus several of the GEROS criteria related to results progression were not applicable.

Exhibit 5.12 Proportion of Reports Reviewed and Good Quality Reports by Results Level (2012)^{11 12}



¹¹ The category activities and products and output were merged for better comparison throughout the years.

¹² Total number of reports per results level: 14 activities & products and outputs, 24 Outcome, and 41 impact.

5.3.5 Stage

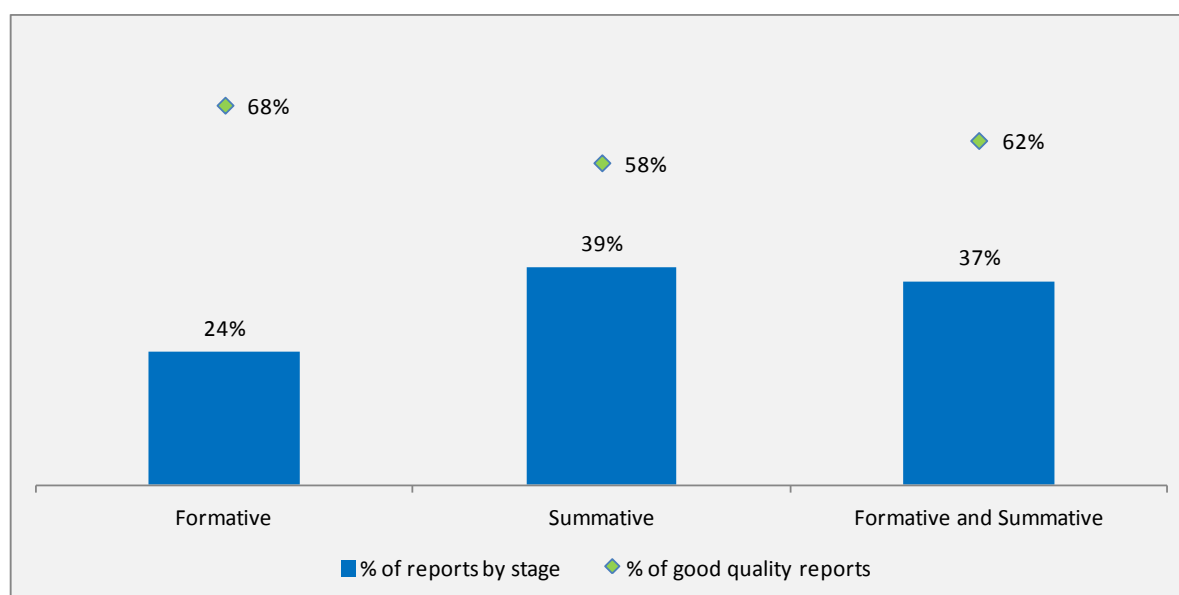
Finding 7: There are no significant trends across evaluations of different stages of an initiative – summative, formative, or a combination thereof. Most evaluation reports have summative or a mix of summative and formative elements.

To better reflect the nature of evaluation reports, an additional category was added to the GEROS template for the review of the 2012 reports. In addition to formative and summative evaluations, close examination of past reports led to the realisation that some reports were addressing both formative and summative considerations. For instance, an end of phase report might discuss progress made within a particular programme, all the while summing-up the contributions made during that specific cycle of investments.

While comparisons across years are not possible due to the change in classification for 2012 reports, there has been a consistent decrease in the volume of summative evaluations since 2009, from 55% to 41% in 2011.

The overall level of quality is relatively even across the evaluations at different stages. Although the differences are not significant from one category to another, it is interesting to note that summative evaluations tended to score lower in the purpose, objectives and scope section with 61% scoring mostly satisfactory and unsatisfactory, as well as the methodology section, where 65% of the reports scored below UNICEF’s standards. By contrast, formative evaluations scored high in every section, with recommendations receiving the worst score (42% of reports were rated mostly satisfactory or unsatisfactory for this section).

Exhibit 5.13 Proportion of Reports Reviewed and Good Quality Reports by Stage¹³



¹³ Total number of reports per stage: 19 formative, 31 summative, 29 formative and summative.

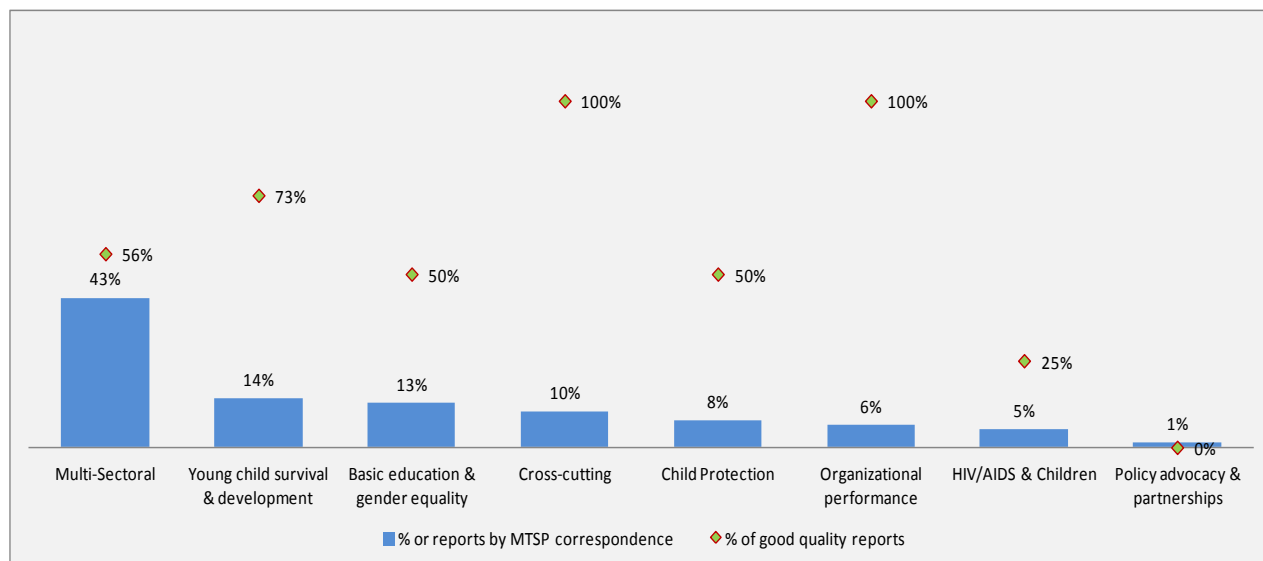
5.3.6 MTSP Correspondence

Finding 8: In 2012, a larger proportion of evaluation reports covered more than two of the MTSP focus areas (known as multi-sector evaluations). In addition, performance levels have increased in most of the focus areas compared to 2011. Policy Advocacy and Partnerships and HIV/AIDS and Children are the two focus area priorities with lowest ratings, but also include fewer evaluation reports.

The trends in the Medium-Term Strategic Plan (MTSP) correspondence are largely in accord with past review cycles. More specifically, the proportion of multi-sector evaluations has increased even further in 2012 such that it accounted for 43% of submitted reports compared with 25% in 2011 and 12% in 2009. The sharpest drop in terms of the number of submissions was in relation to basic education and gender, which fell from a high of 25% in 2011 to only 13% in the 2012 reports. Performance levels have increased across nearly the entire portfolio of reviewed reports. As shown in Exhibit 5.14 below, only studies related to HIV/AIDS and Policy Advocacy failed to meet UNICEF standards by 50% or more, though the low number of reports submitted in these areas invariably heightened the pull of weaker contributions. Compared to 2011, the quality of reviewed reports increased for all categories, except for Policy Advocacy.

Although there is no clear relationship between the type of report and the overall rating, evaluations concerning Young Child Survival and Development tended to be particularly good at explaining the background and object of the evaluation, reporting on the findings and conclusions, and developing the recommendations and lessons learned. A large majority of these reports (82%) were clear, logical and coherent. Although few in number (eight), evaluations reporting on cross-cutting themes, tended to be strong in every aspect, except in describing the methodology. Similarly, evaluations related to organizational performance tended to be good in every aspect, with a slight weakness in reporting on the findings and conclusions.

Exhibit 5.14 Proportion of Reports Reviewed and Good Quality Reports by MTSP Correspondence¹⁴



¹⁴ Total number of reports per MSTP Correspondence: 34 multi-sectoral, 11 young child survival & development, 10 basic education & gender equality, 8 cross-cutting, 6 child protection, 5 organizational performance, 4 HIV/AIDS & children, and 1 policy advocacy & partnerships.

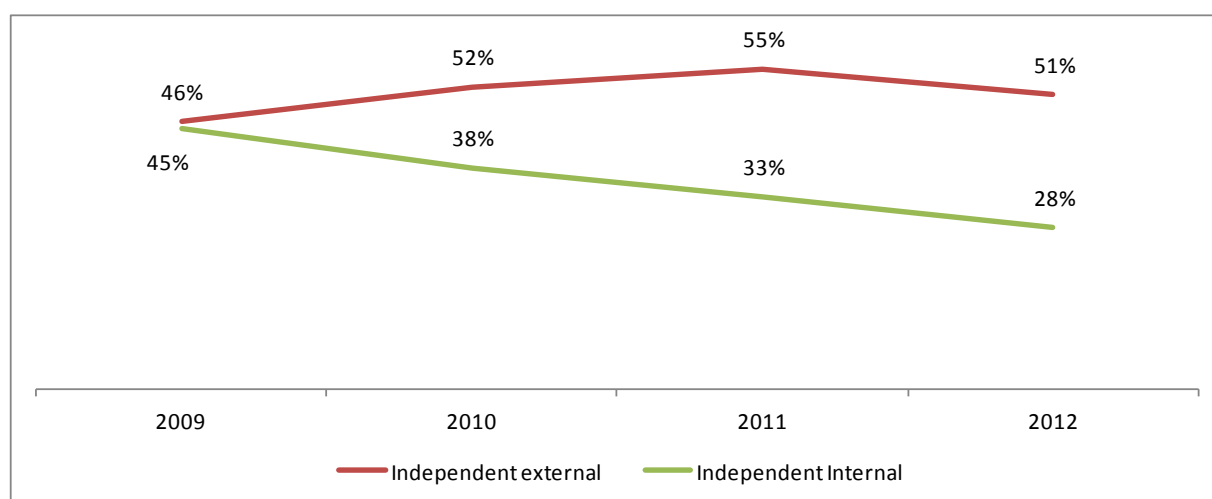
5.3.7 Level of Independence

Finding 9: The move towards externally managed independent evaluations noted in the last meta-evaluation was maintained in 2012, with an even more pronounced shift in favour of independent external¹⁵ reports.

About half (51%) of submitted reports were considered to be independent external, compared with 55% in the preceding cycle of reviews. In comparison, 28% of the reports were noted as being independent internal¹⁶ by nature. The decrease in the number of independent external and independent internal reports for 2012 could have been caused by the larger percentage of reports that could not be catalogued this year (22% rather than 10% in 2011).

The proportion of reports that were rated outstanding and highly satisfactory was greater for independent internal (77%) than for independent external (63%). Given that none of the reports identified as independent internal were rated unsatisfactory and only one was judged to be mostly satisfactory, a more in depth analysis of the relative strengths and weaknesses of internal versus external reports is not possible. Overall however, there appear to be a shift towards independent and externally managed evaluations since 2009, which would tend to increase the impartiality of the report.

Exhibit 5.15 Proportion of Reports by Level of Independence (2009-2012)



5.3.8 Trends by Language of Report

Finding 10: English is still the most common language of UNICEF’s evaluation reports. This year, there are no discernible differences in quality that can be linked to the language of the report.

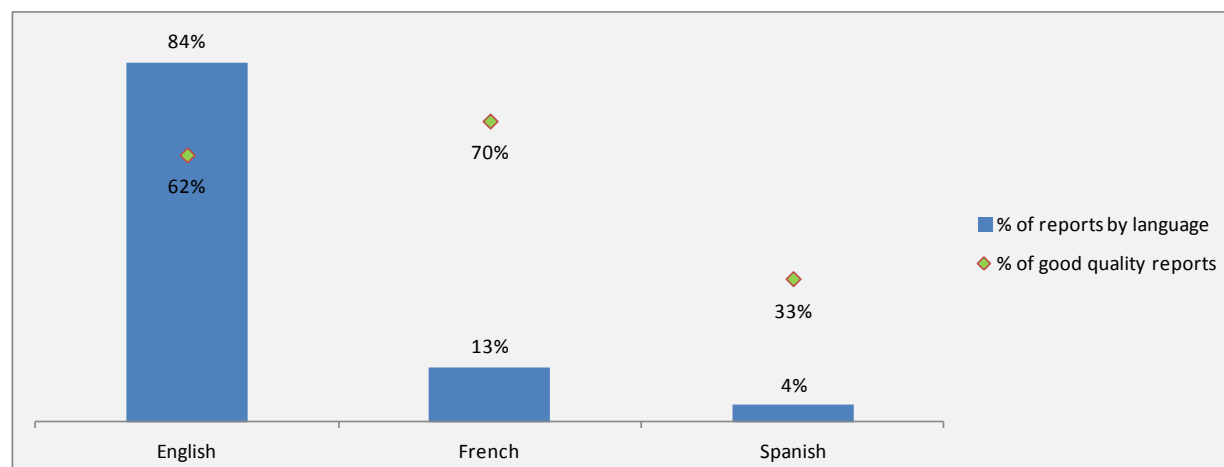
The vast majority of submitted reports were in English, which accounted for more than 83% of reviews. The proportions in number of reports in English, French and Spanish, differs little from past meta-evaluation reports. But unlike in previous years, analysis of the 2012 reports found no discernible differences in terms of the quality of submitted reports based on their language. Thus, 62% of English reports were rated as highly satisfactory or outstanding while 70% of French reports were found to have met UNICEF standards, representing a threefold increase from 2011. As for Spanish reports, the low

¹⁵ “Independent external” means that the evaluation is implemented by external consultants and/or UNICEF Evaluation Office professionals. The overall responsibility for the evaluation lies outside the division whose work is being evaluated.

¹⁶ “Independent internal” means that the evaluation is implemented by consultants but managed in-house by UNICEF professionals. The overall responsibility for the evaluation lies within the division whose work is being evaluated.

number of submissions makes it impossible to discern any real pattern in terms of their overall quality. The 2012 meta-evaluation report speculated that early availability of GEROS guidelines and guidance notes in English probably had something to do with previously observed performance levels in French and Spanish.

Exhibit 5.16 Proportion of Reports Reviewed and Good Quality Reports by Language¹⁷



5.3.9 Coherence

Finding 11: More than half of the evaluation reports (58%) reviewed in 2012 were considered to provide a coherent overall narrative.

For GEROS, coherence refers to the internal logic and consistency in a report. In 2012, 58% of the evaluations reviewed demonstrated overall coherence while 30% were somewhat coherent and 12% were incoherent. For the reports that did not meet all of the standards, the most commonly recurring issues noted were lack of differentiation between findings, conclusions, lessons learned and recommendations (24% of reports), reader-unfriendly formatting (21%), as well as poor quality of and unclear links between individual sections (21% and 17% respectively).

5.4 Trends by Quality Assessment Category

5.4.1 Overall trends Sections A-F

Finding 12: The quality of reports for 2012 increased in all of the sections of the UNICEF-Adapted UNEG standards for evaluation reports.

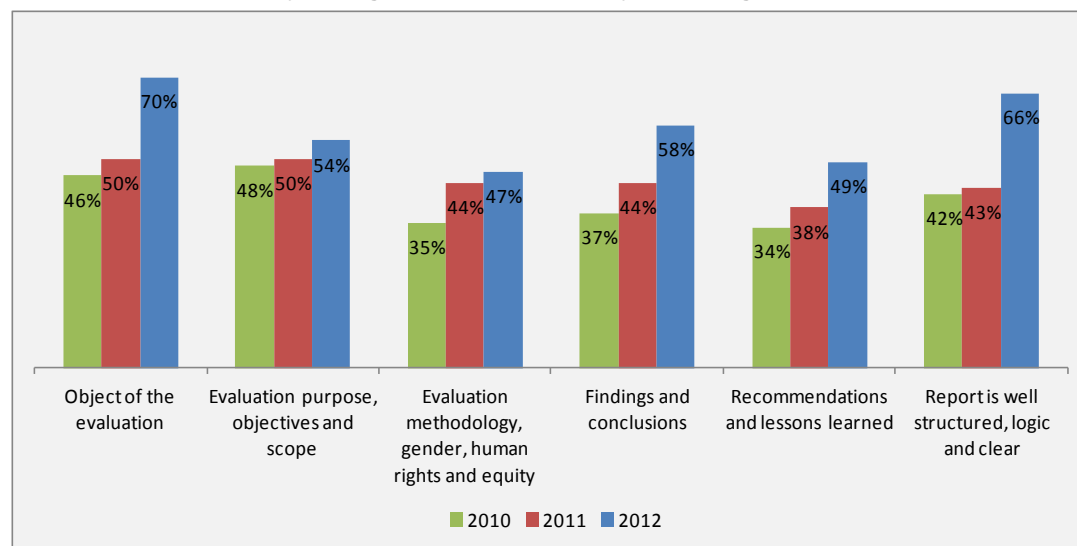
As mentioned earlier, the GEROS assessment is structured around six categories of standards—the description the evaluation object (Section A); the purpose, scope and objectives of the evaluation (Section B); the evaluation methodology and inclusion of human rights, gender and equity (Section C); the findings and conclusion of the report (Section D); the recommendations and lessons learned (Section E); and the report structure, logic and clarity (Section F)

As shown in Exhibit 5.17 below, the positive trend towards increasingly sound and good quality reports was maintained for 2012. Compared with results achieved in 2010 and 2011, the population of reports submitted in 2012 not only did better in terms of meeting and surpassing past achievements, but did so by a wide margin in several distinct areas. With respect to the object of the evaluation, performance ratios

¹⁷ Total number of reports per language: 66 English, 10 French and 3 Spanish.

increased by 20 percentage points to achieve a 70% rate of satisfaction with UNICEF Standards. Good quality ratings for findings and conclusions also increased substantially from 44% in 2011 to 58% in 2012. And the most important proportional increase was in terms of the extent to which reports were well structured, logical and clear, which went from a 43% rate of compliance in 2011 to 66% in 2012. Other increases were less dramatic but no less important. Each of these results were considered individually in the following sub-sections to better understand emerging strengths and areas for further improvement

Exhibit 5.17 Good Quality Ratings per Section- Year by Year Progression (2010-2012)

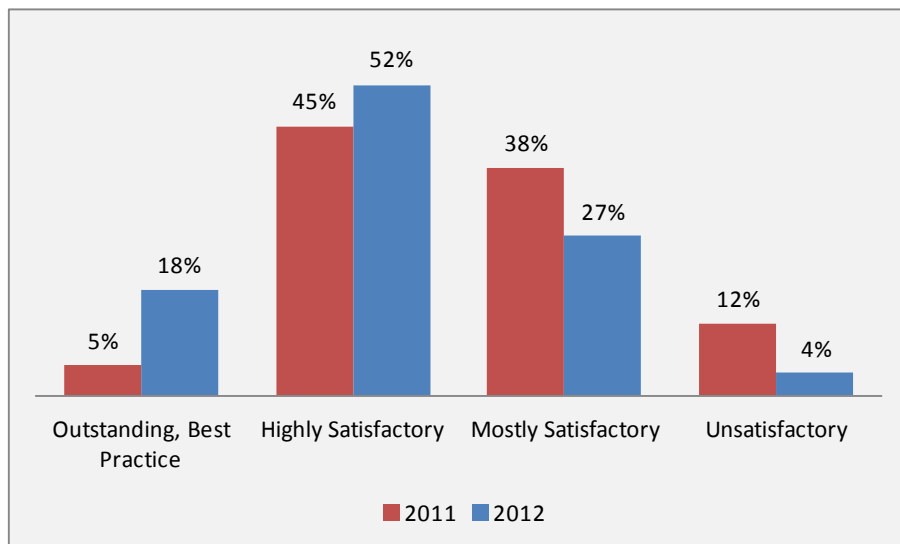


5.4.2 Section A: Object of the Evaluation

Finding 13: Though the extent to which the object of the evaluation is clearly defined has considerably increased, in comparison to previous years, the articulation of a clear theory of change or results logic remains weak overall.

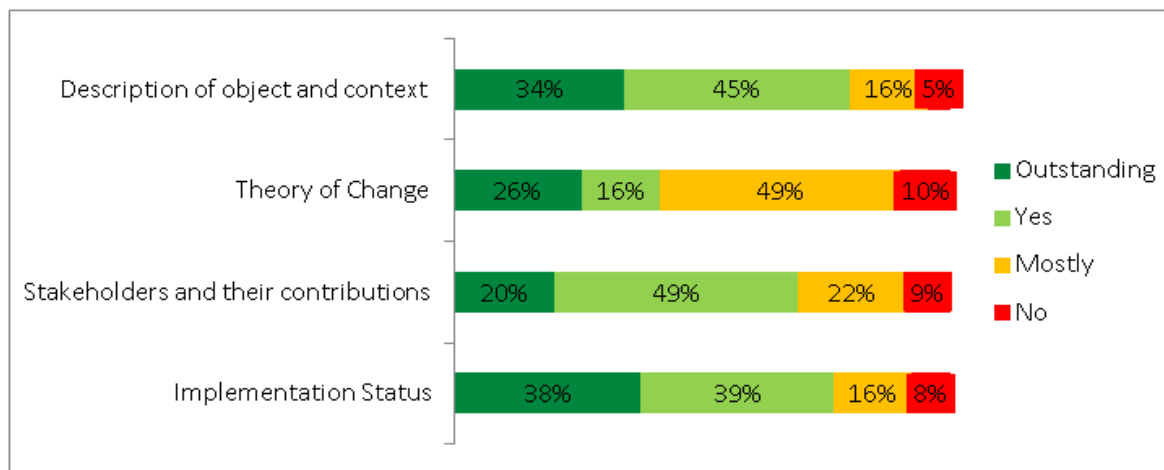
There was a significant improvement in this section from last year with 70% of the reports achieving UNICEF quality standards, compared to 50% for reports submitted in 2011. Moreover, there was an 8 percentage point decrease in the proportion of reports rated unsatisfactory in this section (see Exhibit 5.18). Our analysis of submitted reports suggests that improvements in the quality of submissions are in part attributable to better descriptions of the object and context of the evaluations, for which good quality ratings increased by 20 percentage points between 2011 and 2012. A clear majority (83%) of evaluations reviewed were commended for providing a clear and thorough description of the object evaluated and 77% of the reports were also positively reviewed for including a description of the relevant contextual elements which both justify the need for the intervention and help explain the successes and failures of the intervention being evaluated.

Exhibit 5.18 Section A Ratings and Comparison with 2011



The sub-section on stakeholders and their contributions is another component that saw a significant improvement from last year, increasing from 49% to 69% in the current review. When disaggregated however, analyses of review results suggest that in fact 82% of the reports identified key stakeholders, but that only 52% provided descriptions of their specific contributions in the evaluated interventions themselves. To this end, reports rating outstanding/best practice for this section tended to provide charts or graphs illustrating linkages between stakeholders (e.g., donors, implementing agencies, partners or collaborators, and beneficiaries), an approach that was found to be particularly useful when many stakeholders or groups were involved.

Exhibit 5.19 Sub-Section Ratings: Object of the Evaluation



One sub-section that did not improve from last year was the theory of change. In fact, the main area for improvement identified by GEROS reviewers in 2012 is the need for reports to present an explicit theory of change or a logic model/results chain. According to review ratings, only 42% of evaluations attempted to construct a logic model and for at least a third of the reports (33%), reviewers noted that the evaluation would have been strengthened had there been a clearer outline of the theory of change that underpinned the initiative or a corresponding reference to a results chain or logic model that would have spelled out

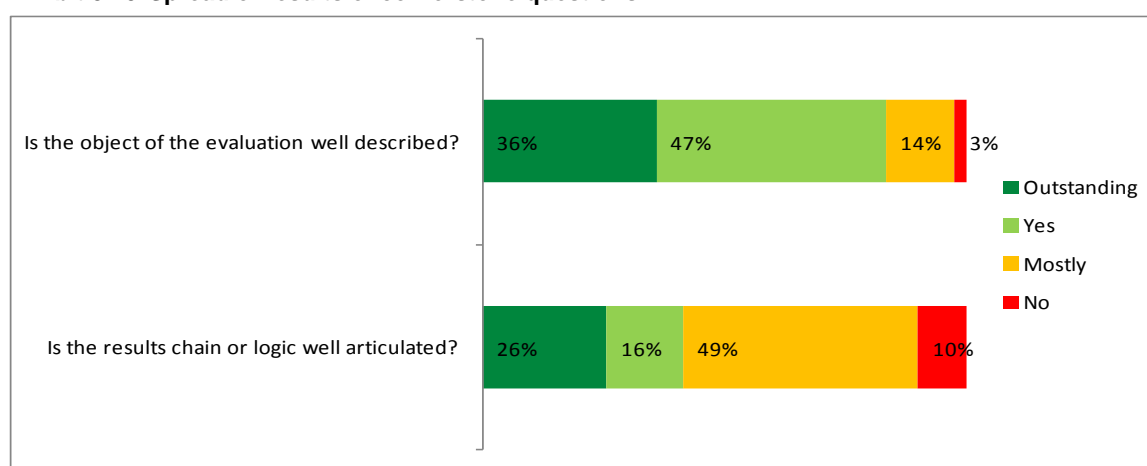
linkages between activities and results at the output, outcome and impact levels (see sidebar and Exhibit 5.20). Although a majority of evaluations were assessed positively for providing contextual information, reviewers also noted that 22% of programming reports did not provide enough contextual information on political, economic, demographic and institutional factors, or other related issues. Hence the unsatisfactory ratings for this section tend to reflect a failure to explain the context of the evaluation as well provide a clear description of the intervention at hand.

Example of review of an Outstanding Section A

“Overall, the object of the evaluation is clearly presented, with the programme’s intended pathway to higher order results clearly identified and the relevance of this framework assessed within the body of the report. Furthermore, the context helps explain choices for the initial programme design and adjustments made during implementation, as well as inform the evaluation findings. Additionally, stakeholders and their contributions are presented in the report.”

Final Evaluation of UNICEF’s “Making PPPP Work for Rural Water Supply in Somalia” (GEROS-Somalia-2012-011)

Exhibit 5.20 Spread of results of cornerstone questions



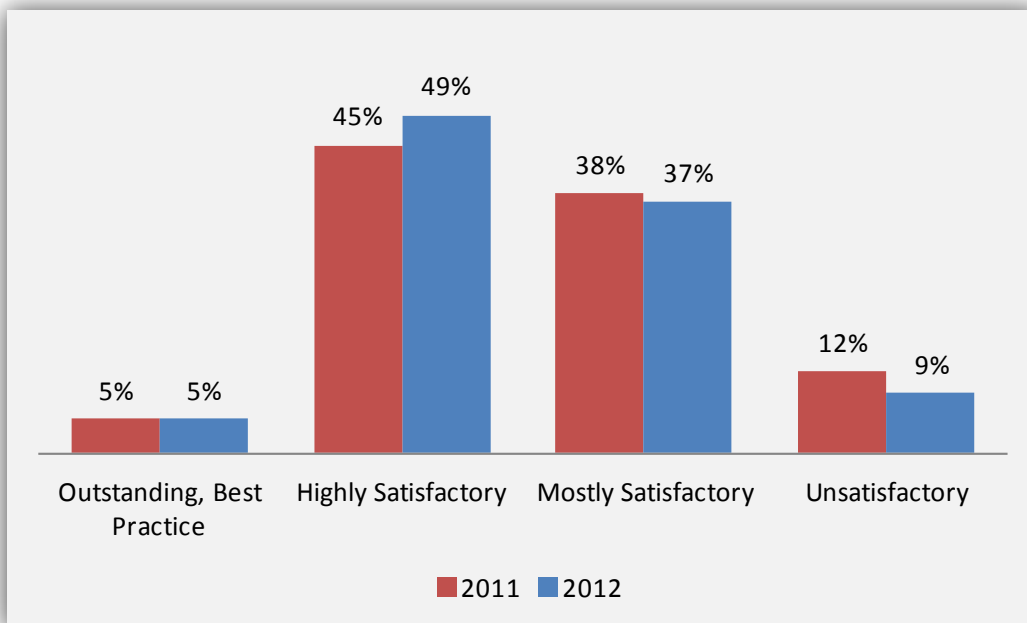
5.4.3 Section B: Evaluation Purpose, Objectives and Scope

Finding 14: The proportion of reports that clearly present the purpose, objectives and scope of an evaluation increased only marginally from the previous cycle of reviews. Failure to specify the underlying questions and criteria guiding an evaluation remains problematic.

Ratings for the section dealing with the purpose, objectives and scope of evaluations, which also includes the evaluation framework, remained relatively stable in comparison to previous cycles of reviews. As shown in Exhibit 5.21, there was only a four percentage point increase in good quality ratings between reports submitted in 2011 and 2012, resulting in a 54% rate of compliance; the proportion of unsatisfactory ratings also experienced a slight decrease, from 12% in 2012 to 9% in 2013.

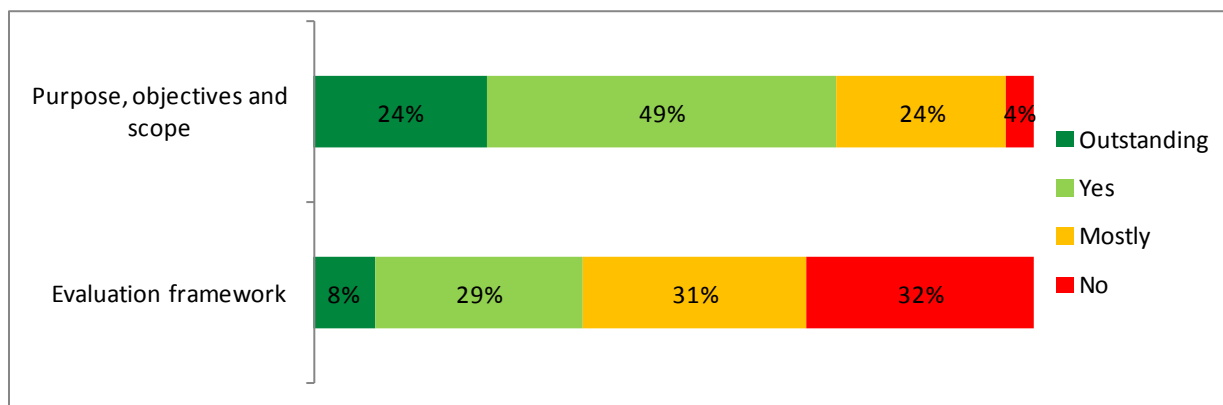
In terms of improvements, the highest improvement in ratings came from the Sub-section “Purpose, objectives and scope”, which rose from 60% to 73% (see Exhibit 5.22). Analysis of report reviews indicates that three quarters (76%) of the reports succeeded in explaining the rationale behind the evaluation and its objectives. Clarity on the scope of the evaluation varied from one report to another, but overall, more than half (57%) of the evaluations succeeded in explaining what will and will not be covered in the evaluation.

Exhibit 5.21 Section B Ratings and Comparison with 2011



While reports were rated as being relatively strong in terms of explaining the purpose, scope and objectives of the evaluation, the majority of reports failed to properly develop or clearly articulate the evaluation framework upon which assessments were based. In fact, good quality ratings for this sub-section experienced a ten percentage point decrease from 47% to 37% between 2011 and 2012. Among other things, qualitative reviews revealed that 77% of the reports failed to provide a justification for selected evaluation criteria. This omission is especially critical for evaluations that rely on criteria other than those recognised by the OECD-DAC, but in cases where standard criteria are used, the extent to which these should be explained or justified is arguably debatable. Exchanges on this issue between the review team and several UNICEF units revealed a lack of consensus on this question and therefore a need to more clearly define the conditionality of this evaluative consideration, with the aim of defining when and how it should apply.

Exhibit 5.22 Sub-Section Ratings: Purpose, objectives and scope of the evaluation



A rating of outstanding/ best practice was given to this section when the purpose, objectives and scope of the evaluation were clear, and an evaluation framework was provided in clear and unequivocal terms. A good practice example would be the description and justification of all of the evaluation criteria used and the inclusion of the corresponding evaluation questions. An unsatisfactory rating in this section tended to reflect a lack of clarity in the purpose, objectives and scope as well as an unclear use or failure to mention the evaluation criteria to be used to guide the assessment. To this end, the review team noted that a handful of reports had failed to use the evaluation criteria correctly altogether. Of related importance to these observations is the fact that the ToRs were provided for only 63% of reviewed reports. This made it harder for reviewers to assert whether and how all evaluation criteria and questions were actually addressed, especially in cases where reports were less than clear on the breadth of the evaluation framework.

Example of review of an Outstanding Section B

“The evaluation purpose, objectives (general and specific) and scope are outlined clearly and concisely. Similarly, the evaluation criteria are listed and described as they relate to the evaluated object.”

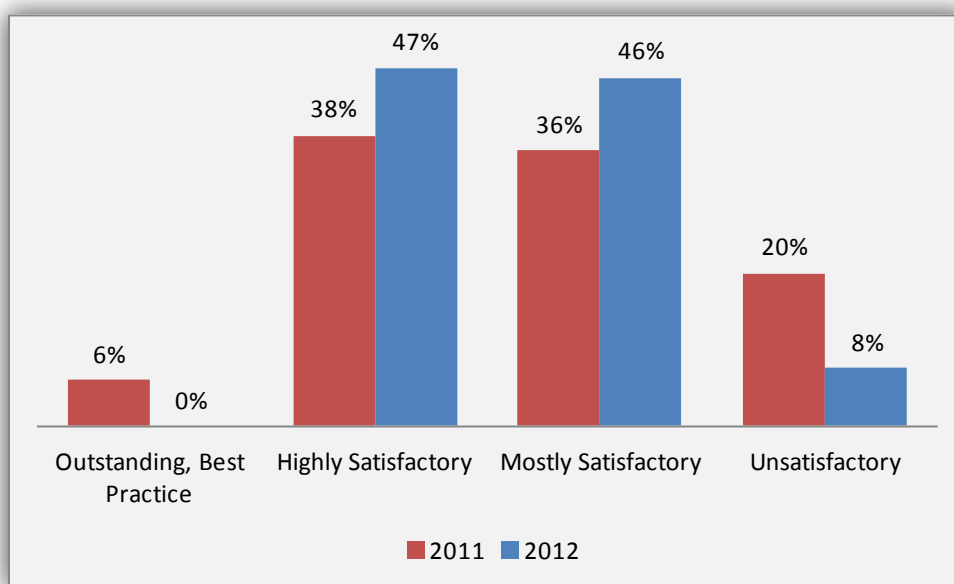
An Evaluation of the Use of Point of Care PIMA CD4 Cell Count Machines for HIV Positive Women and their Families in Maternal Newborn And Child Health (MNCH) Settings in Seven Districts in Zimbabwe, 2012 (GEROS-Zimbabwe-2012-002)

5.4.4 Section C: Evaluation Methodology, Gender, Human Rights and Equity

Finding 15: Compliance with UNICEF standards regarding gender, human rights, equity and methodological considerations continues to progress, though achievements on these issues remain weak, relative to other dimensions.

The overall soundness and appropriateness of methodological considerations improved slightly between 2011 and 2012, with an increase in the rate of conformity to UNICEF standards from 44% to 47% in the submitted reports. Unlike the previous cycle of reviews however, no reports were rated as outstanding or very confident as used in the previously used terminology. Yet, the percentage of unsatisfactory ratings decreased by 12 percentage points, from 20% in 2011 reports to 8% in the latest reviews.

Exhibit 5.23 Section C Ratings and Comparison with 2011

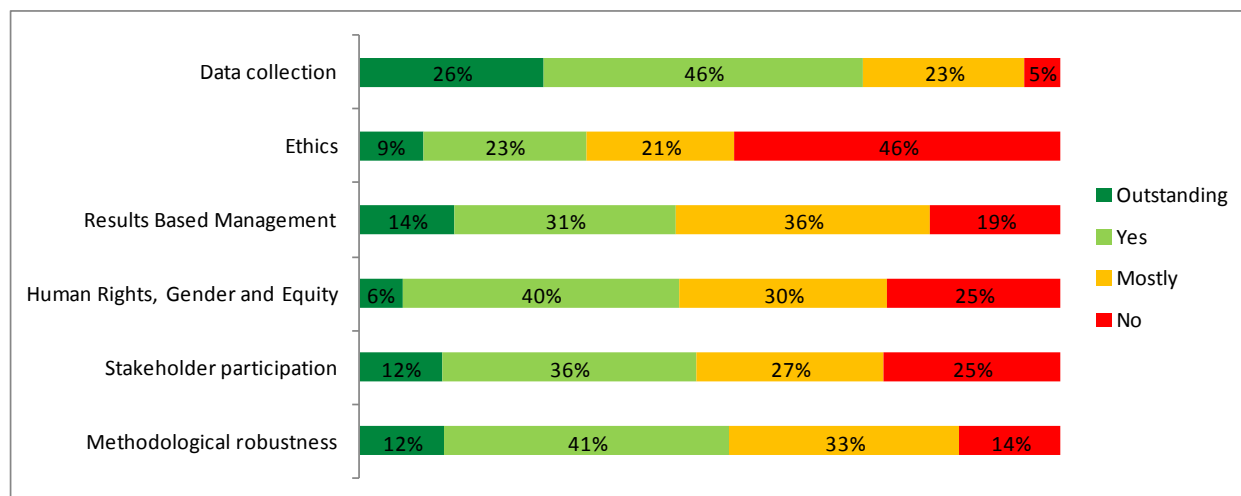


Detailed examination of the various components of this section reveals that 72% of reviewed reports provided clear explanations on the methodological choices that were made regarding the sources of data, sampling strategies, and data collection and analysis. This is a marked increase from the previous year, where only 60% of submitted reports achieved good quality ratings. Though the sub-section on ethics experienced a similar increase in the ratio of reports that achieved good quality, from 22% to 32%, compliance on this issue remains sub-optimal.

Another area where compliance to UNICEF standards increased considerably is that of methodological robustness. Compared to the 2011 submissions, the ratio of reports that achieved high ratings increased from 41% to 53%. Of these, half of the report reviews (24%) stood out for their efforts to triangulate data and control for bias. Generally, such efforts were achieved by taking into account the views of a broad range of stakeholders and the use of mixed data collection methods, such as focus group discussions, interviews, surveys, and document reviews. In short, such reports demonstrated an effort to triangulate and validate the collated sources of data.

With regards to compliance with results-based management principles, 2012 reviews reveal only a slight increase in the number of reports that met UNICEF standards, from 41% for 2011 to 45% in this latest review. In more specific terms, 43% of reports attempted to assess the capability and robustness of the evaluated object’s monitoring system, whereas just over a third (37%) made use of the evaluated object’s M&E or logical framework to guide the assessment. The extent to which reports provided information on the level of participation of stakeholders in the evaluation process (e.g., validation of findings and/or involvement in the development of the recommendations) diminished slightly between submissions for 2011 and 2012, from 52% to 48%. On this point, the reviewers noted that the level of detail provided on stakeholder involvement varied greatly from one report to another, when compliance to such standards could readily be achieved with a minimum amount of effort.

Exhibit 5.24 Sub-section ratings: Methodology, Gender, Human Rights, and Equity



Highly satisfactory methodology sections tended to both clearly present and justify their methodological choices. Quantitative analysis of reviewer comments reveals that only 18% of the reports provided a rationale for the selection of methods regarding data collection, sampling and analysis. Moreover, even though more than half (56%) of the reports were rated favourably in terms of the extent to which they presented a methodology that would facilitate answers to the evaluation questions, only 8% of these provided methodological descriptions that were deemed transparent and explicit by reviewers. To be rated outstanding in this regard, the inclusion of an evaluation matrix was necessary to demonstrate linkages across evaluation criteria, questions, performance indicators, methods and sources of data. Only a handful of reports provided such depth of information, though many cited the existence of similar details in the inception report. Moreover, as discussed above, the vast majority of reports failed to provide a clear and

explicit description of the ethical considerations that were taken into account in the evaluation process, and most importantly, the extent to which efforts were made to ensure the confidentiality of respondents. On this note, reviewers noted that enough information had been provided about quoted individuals so as to compromise their anonymity.

As for the challenges faced by the 8% of reports that were rated as unsatisfactory for Section C, they included: (i) unclear descriptions of the methodology and/or justification of the choices made, (ii) inadequate information about the Human Rights Based Approach, gender equality, and equity, (iii) failure to describe the level of stakeholder participation, (iv) weak descriptions of counterfactual methods when these were needed, (v) unclear or lack of discussions on methodological limitations, and (vi) little to no evidence of efforts to triangulate data and control for bias.

Example of review of a Highly Satisfactory Section C

“The report provides a clear description of the evaluation methodology, inclusive of the data collection methods and sources, corresponding evaluation questions, and justification for the use of the methods of data collection and sampling. Data collection and sampling methods are also well described and justified. There is clear evidence that HRBA, gender equality and equity were taken into consideration throughout the evaluation process. In addition, the evaluation mitigated the weakness of the project M&E system by developing a framework for monitoring the performance of schools that use the CFS approach. This framework builds upon quantitative and qualitative indicators developed by UNICEF and the CFS teams in both project countries for each of the CFS results areas...”

UNICEF Evaluation of the Impact of the Child Friendly School Approach in the Eastern Caribbean (GEROS-Barbados-2012-001)

Human Rights, Gender, and Equity

Another challenging area noted across the reviewed reports concerns the extent to which Human Rights-Based Approaches to Programming (HRBAP), gender equality and equity dimensions were integrated in the methodology, findings, conclusions and recommendations of reviewed reports. The ratio of reports that successfully integrated these elements did not rise above the fifty percentile mark, such that gender was addressed in 46% of the reports, 44% paid attention to HRBAP, and 41% incorporated a greater focus on equity.

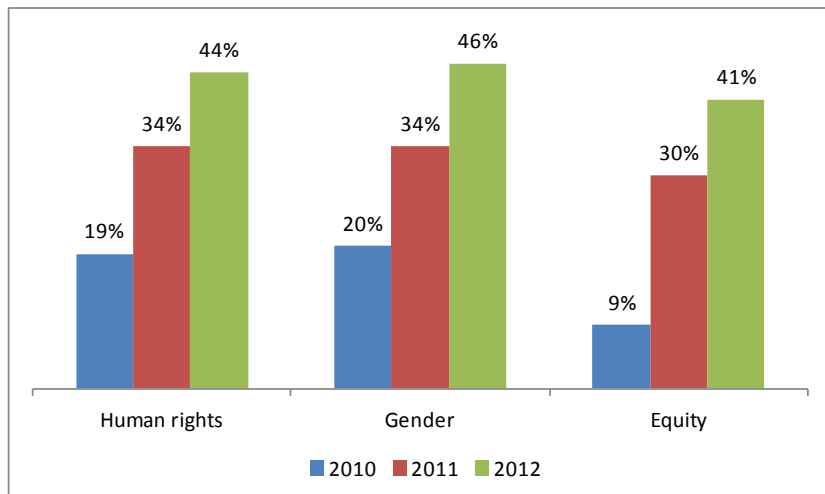
In line with previous years, the trend towards increased integration of human rights, gender and equity concerns remains strong (see Exhibit 5.25). A 13 percentage point improvement in good quality ratings was achieved for the subsection as whole, bringing the total percentage from 33 to 46 percent. Comparisons between Geros review data for 2010-11 and 2011-12 suggests that a greater degree of change was achieved in 2010-11 than in the latter period (see Exhibit 5.25); improvements across the three areas averaged 11% between 2011 and 2012, the increase between 2010 and 2011 averaged 17%. One of the areas that stand out is the integration of equity, which jumped dramatically from 9% to 41% in good quality ratings between 2010 and 2012. This significant increase could be linked to investments made by UNICEF at the beginning of 2011 to strengthen the capacity of Country Offices and partners on how to better integrate equity into the design and management of evaluations (e.g., production of a handbook,¹⁸ webinars, and regional M&E meetings and training events).

With respect to the latest rounds of reviews, analyses of reviewer comments indicates that 41% of the reports had disaggregated data by age, gender, or vulnerability, whereas only 23% used rights-based frameworks, such as the Convention on the Right of the Child (CRC), UNICEF Core Commitments for Children (CCC), and the Committee on the Elimination of Discrimination against Women (CEDAW) to assess the degree to which human rights-based approaches had been duly integrated in the evaluated initiatives themselves. Finally, only 20% of the reports reviewed incorporated human rights language, as well as gender-sensitive and child-sensitive writing, and the extent to which interventions had been monitored through human rights frameworks was discussed in 39% of the reports.

¹⁸ UNICEF Evaluation Office. “How to design and manage Equity-focused evaluations”. 2011

While results are slowly improving, one possible answer for why such considerations are not better reflected in UNICEF-sponsored evaluations probably has to do with the nature of the ToRs themselves and whether or not expectations regarding gender, equity and HRBAP are clearly spelled out. On this note, the review team found that in most instances where these elements are under-represented or poorly addressed, the ToRs also tended to lack clear instructions or specific demands on these issues. Improving the integration of such elements obviously begins by establishing clearer instructions prior to the commencement of the evaluation itself. This observation applies to all of the areas considered in Section C, including ethics, results-based management and stakeholder participation.

Exhibit 5.25 Inclusion of Human Rights, Gender, and Equity: Good Quality Ratings Year by Year

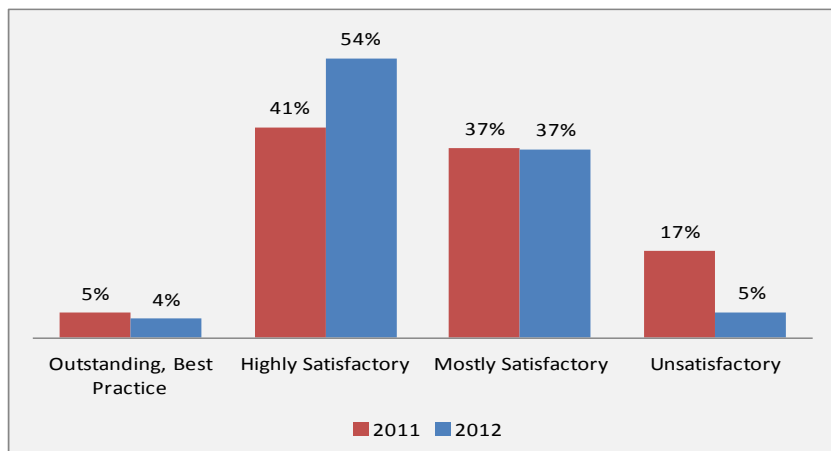


5.4.5 Section D: Findings and Conclusions

Finding 16: UNICEF sponsored evaluation reports have achieved commendable improvements in all areas pertaining to findings and conclusions. Areas for further improvement include the need to better analyse or explain cost-related factors and the development of conclusions that add value to report findings.

There was a marked improvement in the quality of findings and conclusions this year. Good quality ratings for this section increased from 44% to 58%. This positive change is also reflected in the number of unsatisfactory ratings, which decreased from 18% to 5%, and the level of improvements achieved for all sub-section areas.

Exhibit 5.26 Section D Ratings and Comparison with 2011



With respect to the completeness and logic of findings, the ratio of reports adhering to UNICEF standards increased from 50% for 2011 to 65% in the latest set of reviews, and the objective use of evidence was discernible in 71% of the reports. In the best of reports, reviewers noted that findings tended to be clearly and logically organized, and supported by strong and credible evidence. Weaker areas in this sub-section included the general difficulty reviewers faced when trying to assert whether all criteria and questions were addressed, and the fact that gaps and limitations were seldom discussed regarding the availability of the data to answer all of the evaluation questions. As such, only 41% of the reviewed reports attempted to establish clear linkages between findings and the evaluation criteria, either by structuring the report around these criteria or by way of explicit reference in the proposed analysis.

In spite of a 20 percentage point increase in the ratio of good quality ratings, the 44% rate of compliance attributed to cost analysis still makes it the weakest area of Section D. Though this element cannot always be addressed, depending on the type of evaluation and the nature of the evaluated object, reports should at least justify why such an analysis was not or could not be provided. Overall, 26% received a mostly satisfactory rating and 29% were considered unsatisfactory due to the absence of a cost analysis or justification for its exclusion.

On the other hand, the sub-section on contribution and causality tended to fare much better. As one of the strongest areas of Section D, with 68% of positive ratings, reviewers noted that more than three quarters of the reports reviewed (77%) successfully identified causal reasons for accomplishments and failures, though slightly more than half of the reports (57%) attempted to attribute results to specific stakeholders.

Exhibit 5.27 Sub-section ratings: Findings and Conclusions



The strengths, weaknesses, and implications sub-section also improved by 15 percentage points, from a high of 61% for reports from 2011 to 76% in 2012. Specifically, more than two thirds of the reports (69%) presented the strengths and weaknesses of the evaluated object in the conclusions and 77% discussed the future implications of continuing constraints.

Finally, the overall quality of the conclusions in the reports submitted in 2012 rose from 43% in the previous year to a new high of 62%. While a third of the reports (33%) were applauded for presenting conclusions that were based on solid evidence, reviewers noted that in 10% of the reports, conclusions were altogether absent, they were unclear in 19% of the cases, too brief in some instances (17%), or did not add value to findings for 35% of the reports. Hence, according to reviewer comments, only 32% of the reports were noted for the quality of their conclusions and the value they added to the findings. Finally, 58% of the reports were commended for the fact that conclusions were pitched at the right level for end-users and little more than half (54%) appeared to have taken into account the views of all concerned stakeholders.

In order to obtain a rating of outstanding or highly satisfactory, findings first needed to be visually clear and easily identifiable. Reviewers noted that this was not the case in at least 17% of the reports and suggested using different fonts or text boxes to highlight finding statements. In addition, the justification and reasoning behind findings needed to be logical and coherent, as well as supported by strong qualitative and quantitative evidence, wherever possible. As indicated above, this appears to have been done for the majority of reviewed reports (72%). Another element of good practice was to link findings to the evaluation criteria and questions. While evaluation criteria were clearly presented and applied in 41% of the reports – at times, even serving to structure the analysis section – evaluation questions were found lacking or inadequately articulated in 28% of the reports. The generalised absence of succinctly developed evaluation matrices (see Section 5.4.4 above) further hampered clarity in this regard.

Example of review of an Outstanding Section D

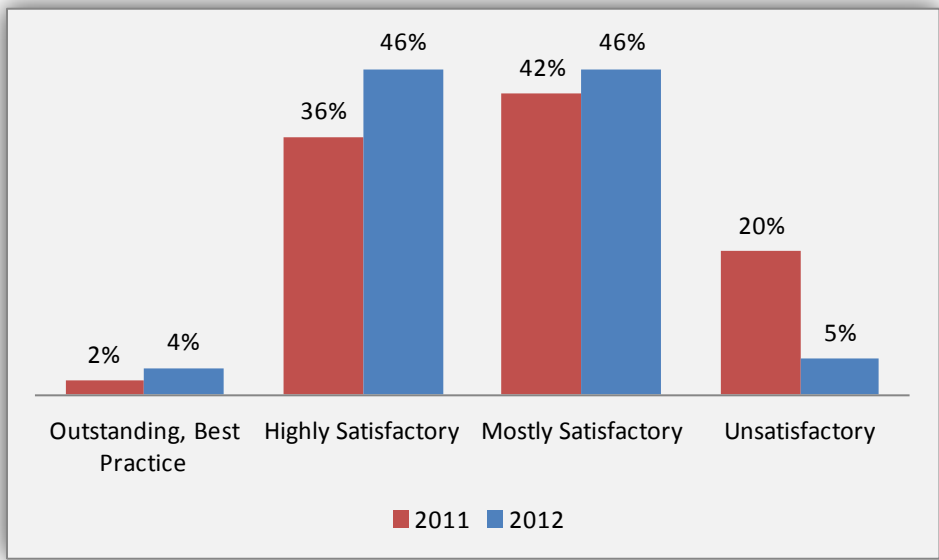
“The report clearly presents the findings and the conclusions of the evaluation. The structure of the report facilitates the presentation of the findings, as a section is dedicated to each evaluation criterion. The evaluation questions are covered and supported by sufficient evidence and are reiterated in the beginning of each section. The conclusions provide relevant insight into the object and the purpose of the evaluation and are pitched at an appropriate level.”

Global Evaluation of Life Skills Education Programmes (GEROS-USA-2012-011)

5.4.6 Section E: Recommendations and Lessons Learned

Finding 17: Though improvements in the sections on recommendations and lessons learned are accruing at a more modest pace, evidence suggests that substantial gains in these areas could be achieved by clearly differentiating lessons learned from the conclusions and recommendations and by improving the usefulness of recommendations.

Exhibit 5.28 Section E Ratings

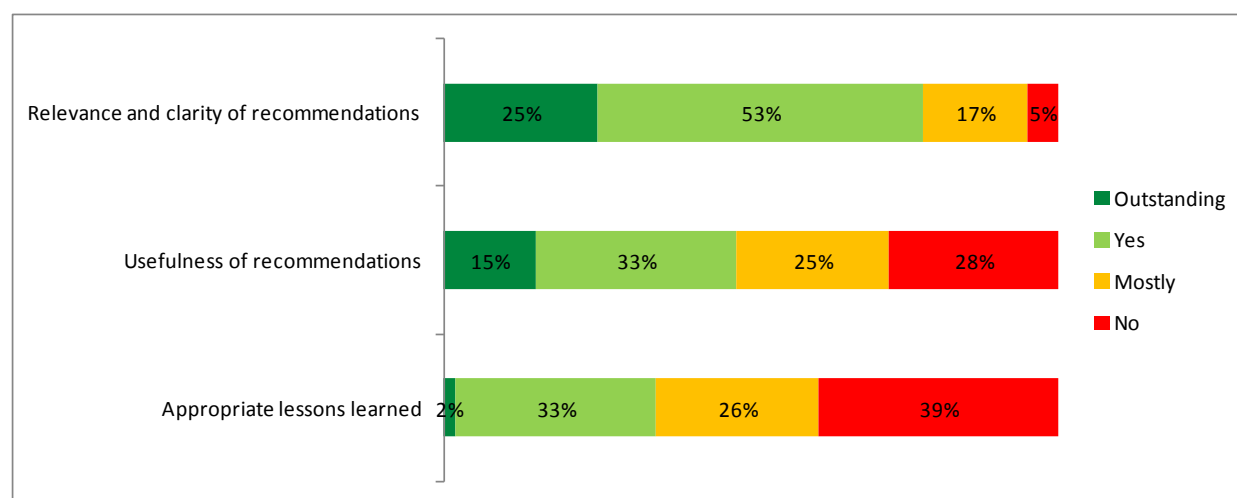


Despite being one of the weakest sections of the reports submitted in 2012, notable improvements were achieved regarding the presentation of recommendations and lessons learned since the previous cycle of reviews. Whereas report ratings for this section showed a 38% rate of compliance with UNICEF standards for submissions made in 2011, half of all reviewed reports from 2012 were rated highly satisfactory or outstanding, and a 15 percentage point drop in the number of that were considered unsatisfactory was likewise registered. Further, according to results of reports from 2011, only 56% were

rated as highly satisfactory for their relevance and clarity, compared with 78% for 2012 reports. Even more striking was the 22 percentage point increase in outstanding ratings for this sub-section. Though positive overall, the major weakness in this sub-section related to the extent to which recommendations were clearly stated and prioritised. Only 57% of the reports achieved a passing mark on this question.

There was also a slight improvement in the usefulness of recommendations, from 40% in 2011 to 47% in 2012. For the most part, reports tended to clearly identify the target group for action in 65% of the reports and provided recommendations that appeared to be realistic in the context of the evaluation in 58% of reviewed cases. The most striking omission in evaluation reports was the lack of a description of the process followed in developing the recommendations. Only 19% of the reports managed to address this issue correctly.

Exhibit 5.29 Sub-Section Ratings: Recommendations and Lessons Learned



The strongest area in this section concerned the relevance and clarity of the proposed recommendations. Overall, our analysis reveals that 86% of the reports received ratings above highly satisfactory for clearly developing recommendations on the basis of the data presented in the findings and the conclusions. Moreover, 89% of these were deemed relevant to the object and purpose of the evaluation. As mentioned above, areas that showed the weakest improvements centered on the usefulness of recommendations and the appropriateness of lessons learned. For instance, nearly three quarters of the reports (64%) would have received better ratings if they had simply explained whether and how stakeholders were involved in the development of the recommendations. In order to increase the usefulness of evaluations, reviewers commented in 38% of the reports that recommendations ought to be prioritized and that responsibility for implementation should be attributed to specific stakeholder group in 31% of the reports. The sub-section on lessons learned was the weakest area of this section overall, with 32% of the reports either not including such a section or failing to distinguish it from the conclusions and recommendations, and another 26% that did so only partially. Moreover, in only 27% of the reports did reviewers find lessons learned that could be generalized to other contexts. In general, lessons learned, when provided, tended to be too project-specific.

Examples of review of an Outstanding Section E

« Les recommandations donnent non seulement des conseils sur les éléments du programme qui doivent être améliorés, mais dressent aussi la liste des actions spécifiques qui doivent être entreprises pour la mise en œuvre de celles-ci. »

« De plus, les recommandations contiennent une dimension de responsabilité puisqu'elles sont mises en correspondance avec un partenaire spécifique pour faciliter le processus. »

« Les leçons apprises résumant bien les points clés de l'expérience du programme au Tchad. »

Évaluation de la Prise en Charge Communautaire de la Malnutrition Aiguë (PCMA): Étude de Cas du Tchad (GEROS-Chad-2012-004)

5.4.7 Section F: Report structure, logic and clarity

Finding 18: Overall, evaluation reports are clearly and logically structured, and executive summaries tend to be relatively complete, though more could be done to ensure that executive summaries can stand alone.

While the past year’s ratings for this section had been fairly consistent across years, this year’s quantitative analysis points to a significant improvement from last year’s review, with 65% of the reports rated highly satisfactory or outstanding compared to only 43% in 2011. There was also a 5 percentage point decrease in the reports rated unsatisfactory/not confident to act. Significant improvements were noted for both sub-sections on style and presentation and the executive summary. Indeed, the proportion of outstanding reports for style and presentation increased from 4% in 2012 to 20% in 2013, while an 8 percentage point drop was observed for unsatisfactory reports. Moreover, reports that received an outstanding rating for its executive summary jumped from 3% in 2011 to 25% in 2012 and unsatisfactory reports, standing at 13% in 2011, reached a record low of 1%.

Exhibit 5.30 Section F Ratings and Comparison with 2011

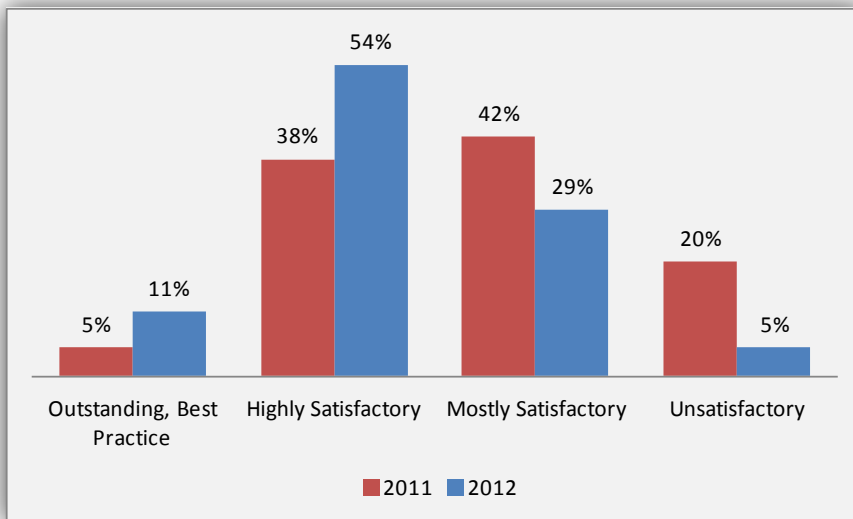


Exhibit 5.31 Sub-Section Ratings: Report Structure, Logic and Clarity



Of the two sub-sections assessed, the sub-section on the executive summary was deemed to be the strongest by both the quantitative and qualitative analysis, with more than three quarters (78%) of the reports rated highly satisfactory or outstanding. Most reports (95%) included an executive summary, which represents a slight increase (2 percentage points) from 2011. The extent to which these contained all the necessary elements was noted in 70% of the reports (up from 57% in 2011), and 68% were considered sufficiently complete to stand alone. However, only 58% of the executive summaries were believed to be sufficient to inform decision-making. In terms of potential improvements, 18% of the qualitative reviews suggested that executive summaries could be shorter and thus more concise. Summaries that were 8 to 10 pages long were not uncommon.

By way of comparison, 68% of reports reviewed were rated highly satisfactory or outstanding for style and presentation, compared to only 52% in 2011. Specifically, 70% of the reports were regarded as being logically structured, though 13% of the reviews found that findings should be structured more around the evaluation criteria rather than around the evaluation questions.

Moreover, 16% of the report reviews suggested that a better layout (i.e. use of bolds, graphics, tables) could improve the structure of the report by enhancing the visibility of different sections and findings, and that editing for grammatical, punctuation, orthographic, and typographic errors could enhance the readability of the reports. Finally, less than half (46%) of the reports included annexes containing all the relevant elements, with most of the reports lacking either the Terms of References, methodology, interview protocols, or evaluation matrices.

Example of review of an Outstanding Section F

“Overall, the evaluation report demonstrates a clear style and presentation. The structure has been built with clarity and logic, which facilitates comprehension. The first pages of the report contain all the necessary elements to present the evaluation and also include an outstanding executive summary.”

“It was an excellent decision to include an executive summary in different languages. Future reports at the global level should follow this practice. This would help to reach a greater audience and inform a larger number of decision-makers.”

Global Evaluation of Life Skills Education Programmes (GEROS-USA-201-011)

5.5 Examples of Good Practices

Information from the reviews was analysed in order to extract examples of good practices in different evaluation reports. The table below highlights the main good practices retrieved:

<p>Section A: Object of the Evaluation</p>	<ul style="list-style-type: none"> • Demonstrating expertise in the sector in which the evaluation is concerned makes the analysis more rigorous and robust and it helps with the understanding of how the context affects the evaluation object. (GEROS-USA-2012-008) • A graphic as well as written description of the theory of change clarifies what the object of the evaluation was meant to achieve. (GEROS-Bulgaria-2012-002) • Reconstructing the objectives and logical framework of the evaluation object in consultation with the programme stakeholders when they are unclear ensures an accurate comparison of achieved versus expected results. (GEROS-Sudan-2012-009)
<p>Section B: Evaluation purpose, objectives and scope</p>	<ul style="list-style-type: none"> • A detailed description and justification of the evaluation criteria used clarifies how the evaluation criteria will be used in the evaluation. (GEROS-USA-2012-003)
<p>Section C: Evaluation Methodology and Gender, Human Rights and Equity</p>	<ul style="list-style-type: none"> • Seeking inputs from a wide range of stakeholders, including the beneficiaries, gives the analysis credibility and reflects efforts to control bias. (GEROS-USA-2012-005) • Refining findings and conclusions through a participatory approach by holding a workshop with key stakeholders ensures accuracy in the report and ownership of stakeholders. (GEROS-USA-2012-007) • Exposing factors such as missing data and incomplete indicators from the programme as limitations explains why some evaluation questions or issues were not addressed. • Inclusion of an evaluation matrix provides clarity on the methodological approach used to address each evaluation question. (GEROS-Macedonia-2012-008)

	<ul style="list-style-type: none"> When conducting an impact evaluation, a counterfactual specifically tailored for the programme in question is used to counteract the limitations posed by other traditional methods of establishing counterfactuals. (GEROS-South Africa-2012-001)
Section D: Findings and Conclusions	<ul style="list-style-type: none"> Including anecdotes and success stories is a good way to present evidence to support the findings. (GEROS-Democratic Republic of Congo-2012-001) The incorporation of respondent perceptions through quotes enhances the presentation of the findings. (GEROS-Botswana-2012-004) Explicitly linking results of the programme to its expected outcomes underscores the level of success or failure of the programme. (GEROS-China-2012-008) The evaluation findings and conclusions present a balanced view of the stronger and weaker aspects of the evaluated object. (GEROS-Moldova-2012-003)
Section E: Recommendations and Lessons Learned	<ul style="list-style-type: none"> Writing the recommendations in a structured table, indicating the timelines, feasibility, and target groups helps the users of the evaluation in implementing them. (GEROS-Senegal-2012-008) Grouping recommendations in categories (e.g. operational, strategic, programming), enhances the way recommendations are structured. (GEROS-India-2012-011) A SWOT analysis in the recommendations creates opportunities for continued discussion and increased appropriation by stakeholders. (GEROS-Moldova-2012-002) Cross-referencing recommendations with the findings helps the reader to understand where the data for the recommendations comes from. (GEROS-USA-2012-014)
Section F: Report is well structured logic and clear	<ul style="list-style-type: none"> Structuring the report's findings around the evaluation questions makes the report more useful to the evaluation's users. (GEROS-Liberia-2012-003) For evaluations at the global level, an executive summary in the three official languages (English, Spanish, and French) helps to reach a greater audience and inform a larger number of decision-makers. (GEROS-USA-2012-011)

6. Conclusions

These conclusions are based on the team's analysis and multiple discussions on the main issues cutting across the findings of the 2012 meta-evaluation.

Over half of reports – 62% – meet UNICEF's quality standards (with a rating of highly satisfactory or outstanding), a 20 percentage point increase from 2011 (42%).

Evidence drawn from this and previous review cycles shows an unmistakable trend towards improved report submissions. Overall, the ratio of good quality reports increased significantly from 2011, the largest improvement in report quality since UNICEF initiated GEROS. The proportion of unsatisfactory reports actually diminished from 23% to 8% during the same period. Repeated across the entire breadth of issues addressed in the GEROS framework, higher levels of achievement were registered for five out of the eight regions, and nearly no change was registered for EAPRO, ESARO and TACRO.

Although there was some variation in ratings across different report types, the extent to which a report satisfactorily met UNICEF standards for high quality reports is not linked to the nature or focus of the report.

There was no relationship found between the quality of submitted reports and their geographic scope (e.g. regional, national or sub-national), purpose (e.g. programme, pilots), MTSP area (e.g. multi-sectoral, basic education), or level of results (e.g. output, outcome, or impact) of an evaluation. Therefore, the quality of the report is not affected by the type of geographic scope, purpose, or MSTP Area it focuses on, but by its ability to successfully adhere to UNICEF quality standards.

There was some variation between report sections in terms of overall quality, with the section on the methodology being the weakest. Human rights, gender and equity continue to improve year after year.

The quality of the reports reviewed did not vary dramatically across sections, although there were some sections that came out stronger than others. Though the methodology section improved when compared to last year, this was noted to be the weakest overall. Ethics is a particular area that was found to be weak, though it has improved since 2011 with a 10 percentage point increase in good quality ratings. The section on the object of the evaluation came out the strongest, with 70% of the reports adhering to UNICEF's quality standards. Human rights, gender and equity have been improving since 2010. Although there is relatively little variation between the three areas with regards to their rating levels this year, the integration of equity into evaluation reports has increased dramatically since 2010, up by 32 percentage points, thus demonstrating a possible outcome of UNICEF's efforts to strengthen their internal capacity in equity.

Evaluation reports tend to follow the Terms of Reference; thus the Terms of Reference must reflect the agency's priorities if the evaluation is to reflect those priorities.

Evaluation reports tend to reflect what evaluators were asked to consider, including the suggested methods, evaluation criteria, data sources, and questions underscored in the ToRs. As such, even though gender, human rights and equity concerns may be central to the values of the Agency, whether or not these are addressed in the evaluation reports tends to be dictated by the ToRs. By the same token, the extent to which ethics, M&E frameworks, contribution and attribution, evaluation criteria, cost analysis, the presence or absence of conclusions and lessons, and the contents of the executive summary and report are clearly discussed or presented in a report largely depends on the guidance given to evaluators.

7. Recommendations

The goal of these recommendations is to facilitate the continuous improvement of the quality of UNICEF's evaluations by both focusing on GEROS as a quality assurance system as well as on broader evaluation improvement within the organisation. While evaluation quality is steadily improving, as reported through GEROS meta-evaluations, there have been concerns about the effects of evaluation coverage. We note that steps are being taken (such as the development of a new Key Performance Indicator) to monitor the challenge of evaluation coverage that is identified in the most recent annual report on the evaluation function submitted to the UNICEF Executive Board.¹⁹

The recommendations have been developed through iterative review of the findings and discussions among the GEROS review team members.

Recommendation 1: UNICEF should continue to systematically communicate the GEROS results as part of its effort to incentivize managers regarding the system, as well as communicate the specific criteria of GEROS to evaluators.

UNICEF uses several means of communicating the GEROS findings to its staff including through regional dashboards and through the intranet; these efforts should be continued and enhanced, particularly in regions where participation is low.

To maximize the effectiveness of the GEROS system, greater efforts should be made to systematically communicate to evaluators the exact quality standards that are used for rating evaluation reports. In this regard, the Terms of Reference could communicate that the evaluation report will be rated by GEROS, and guidance on the GEROS criteria should be provided (either through a website link or a list of the actual GEROS criteria). A focus on including such information in the ToRs may also help ensure that joint evaluations and other non-UNICEF managed evaluations also meet the UNICEF quality standards.

This recommendation is relevant to the Evaluation Office and the Regional Offices.

¹⁹ The report (19 April 2013) points to the decline in the number of evaluations being submitted to GEROS, which has dropped from 140 in 2010 to 99 in 2011 and 79 in 2012. It notes that 38 country offices appear to have conducted no evaluations whatsoever over the period 2009-2011.

Recommendation 2: UNICEF’s internal learning systems around evaluation should continue to be strengthened; the GEROS system can play a role in informing the continuous improvement of that learning system.

The UNICEF Evaluation Office has a strong emphasis on learning and has used multiple approaches to foster learning on evaluation and help build evaluation capacity in regions. One thing that appears to be working well is the evaluation support (such as the Regional Facility) provided by Regional Office M&E advisors to Country Offices, since such support is developing processes to help improve the pool of qualified evaluators, as well as the quality of Terms of Reference, evaluation workplans, and draft reports.

The GEROS meta-evaluation provides insights with regard to certain topics that may require additional attention in such efforts from both headquarters and regions. Guidance and learning opportunities are required both for evaluation managers within UNICEF and for the external evaluators, in particular with regard to:

- Key UNICEF priorities such as HRBAP, gender equality and equity, which need to be adequately reflected in evaluation design, implementation, and final reports.
- As noted earlier, Terms of Reference are key determinants of content of evaluation reports and often key criteria are missing from the Terms of Reference themselves and therefore excluded from the evaluator’s analysis or seen to be outside the scope of the evaluation.
- Certain methodological issues should also be focused upon. For example, ethics continues to be a weak area of the evaluation reports. Evaluation matrices are a crucial tool to make stronger linkages between methodology and evaluation questions (which is often a poorly rated criteria in GEROS) and thus the inclusion of matrices should be encouraged in evaluation workplans.
- Appropriate guidance should be given to evaluation managers on how to ensure that their evaluations are properly resourced in terms of the allocation of person days to particular tasks (especially to analysis and drafting of reports, which are often under-resourced). Guidance would include the preparation of a realistic level of effort that reflects the available budget.

This recommendation is relevant to the Evaluation Office and the Regional Offices.

Recommendation 3: UNICEF should continue to review and continually improve the standards used in the GEROS process, even if this risks compromising comparability of GEROS data from year to year.

UNICEF reviews the GEROS criteria (template) on a periodic basis; this is a good practice that should be strengthened so that the review occurs on a more regular basis. In the next review of the GEROS criteria, the Evaluation Office should consider explicitly identifying and clarifying criteria that are ambiguous and/or that may not be applicable for all types of evaluations (such as humanitarian evaluations) and reviewing issues of scaling that emerge with the change in wording of the rating scale.

This recommendation is relevant to the Evaluation Office.

8. Lessons Learned

According to UNICEF, lessons learned are contributions to general knowledge that refine or add to common understanding, but should not be merely a repetition of common knowledge.

The general characteristics of a strong evaluation report include clearly and directly addressing the evaluation criteria, good structure, and logical linkages threaded throughout; thus while content is important, the presentation of that content is just as important.

The content of an evaluation report is the most intuitively obvious basis for creating a strong report; however, the GEROS process made it clear that *how* information is organized and presented is also a key factor in whether a report achieves UNICEF’s quality standards. This is particularly the case because it is

crucial to address the myriad criteria for a good report while simultaneously ensuring that these are threaded together in a coherent manner. Indeed, a good report must move between theoretical elements and data that supports and gives life to those elements, between achievements (outputs) and results (outcomes and impacts) without losing track of the causal variables that affect performance.

Monitoring the quality of evaluations through a GEROS-type system improves the quality of evaluations.

There has been consistent improvement in the quality of UNICEF's evaluations over the last four years, and 2012 was no exception. This underscores that such monitoring of the evaluation report quality is an important element to ensure better evaluation reports. GEROS has established high standards for UNICEF-sponsored, supported or managed evaluations, and supports the strategic interests of UNICEF by improving the quality, reliability and validity of the information that in turn support stronger programming.

The more that UNICEF makes clear to evaluators the priorities and foci of its evaluation system, the more likely it is that evaluation reports will meet those standards.

Evaluators usually attempt to satisfy the Terms of Reference and thus to adhere to UNICEF's evaluation standards. If they are aware that they will be judged according to the GEROS standards and if they are aware of what they specify, they will strive to meet the standards and thereby produce better evaluation reports.

Strong evaluation reports depend upon appropriate time being allocated to analysis and writing.

Based upon the GEROS experience as well as Universalialia's over 30 years of evaluation experience, strong evaluation reports require time to gather the right data to answer the questions but, more importantly, there is a need for a significant amount of time for analysis of that data and report writing. Allocating appropriate time for these latter stages, giving consideration to the evaluation's size, complexity and scope, will tend to produce better reports.

Appendix I List of Findings

- Finding 1: The quality of reports submitted to UNICEF increased sharply between 2011 and 2012.
- Finding 2: With few exceptions, the quality of UNICEF sponsored evaluations has improved across most of the Regional Offices when compared with previous years.
- Finding 3: The geographic scope of most evaluations continues to be at the national or sub-national level. However, the percentage of good quality reports has proportionally increased for all of the categories of geographic scope.
- Finding 4: The quality of UNICEF managed or co-managed evaluations has increased markedly since 2011.
- Finding 5: While number of evaluations by purpose – programme, policy, humanitarian, country evaluations, etc. – appear to differ little from year to year, the quality of reviewed submissions has improved for nearly all types of evaluation reports.
- Finding 6: The proportion of evaluation reports considering results at outcome and impact level has fluctuated over time. Reports that focus on output-level results, which are fewer in number and are often humanitarian or real-time evaluations, demonstrate the greatest improvement in terms of quality. Reports that address outcomes continue to under-perform with regard to meeting UNICEF quality standards.
- Finding 7: There are no significant trends across evaluations of different stages of an initiative – summative, formative, or a combination thereof. Most evaluation reports have summative or a mix of summative and formative elements.
- Finding 8: In 2012, a larger proportion of evaluation reports covered more than two of the MTSP focus areas (known as multi-sector evaluations). In addition, performance levels have increased in most of the focus areas compared to 2011. Policy Advocacy and Partnerships and HIV/AIDS and Children are the two focus area priorities with lowest ratings, but also include fewer evaluation reports.
- Finding 9: The move towards externally managed independent evaluations noted in the last meta-evaluation was maintained in 2012, with an even more pronounced shift in favour of *independent external* reports.
- Finding 10: English is still the most common language of UNICEF’s evaluation reports. This year, there are no discernible differences in quality that can be linked to the language of the report.
- Finding 11: More than half of the evaluation reports (58%) reviewed in 2012 were considered to provide a coherent overall narrative.
- Finding 12: The quality of reports for 2012 increased in all of the sections of the UNICEF-Adapted UNEG standards for evaluation reports.
- Finding 13: Though the extent to which the object of the evaluation is clearly defined has considerably increased, in comparison to previous years, the articulation of a clear theory of change or results logic remains weak overall.

- Finding 14: The proportion of reports that clearly present the purpose, objectives and scope of an evaluation increased only marginally from the previous cycle of reviews. Failure to specify the underlying questions and criteria guiding an evaluation remains problematic.
- Finding 15: Compliance with UNICEF standards regarding gender, human rights, equity and methodological considerations continues to progress, though achievements on these issues remain weak, relative to other dimensions.
- Finding 16: UNICEF sponsored evaluation reports have achieved commendable improvements in all areas pertaining to findings and conclusions. Areas for further improvement include the need to better analyse or explain cost-related factors and the development of conclusions that add value to report findings.
- Finding 17: Though improvements in the sections on recommendations and lessons learned are accruing at a more modest pace, evidence suggests that substantial gains in these areas could be achieved by clearly differentiating lessons learned from the conclusions and recommendations and by improving the usefulness of recommendations.
- Finding 18: Overall, evaluation reports are clearly and logically structured, and executive summaries tend to be relatively complete, though more could be done to ensure that executive summaries can stand alone.

Appendix II List of Recommendations

Recommendation 1: UNICEF should continue to systematically communicate the GEROS results as part of its effort to incentivize managers regarding the system, as well as communicate the specific criteria of GEROS to evaluators.

Recommendation 2: UNICEF's internal learning systems around evaluation should continue to be strengthened; the GEROS system can play a role in informing the continuous improvement of that learning system.

Recommendation 3: UNICEF should continue to review and continually improve the standards used in the GEROS process, even if this risks compromising comparability of GEROS data from year to year.

Appendix III Terms of References

TERMS OF REFERENCE FOR GLOBAL EVALUATION QUALITY OVERSIGHT SYSTEM EVALUATION OFFICE

Background

UNICEF put in place an Evaluation Quality Assurance System to ensure evaluations managed/supported by UNICEF meet quality standards. The system is composed of a) the **Global Evaluation Reports Oversight System (GEROS)**, managed by the Evaluation Office (EO), which rates final evaluation reports commissioned by UNICEF Country Offices (CO), Regional Offices and HQ divisions against the UNEG/UNICEF Evaluation Report Standards, and b) **Regional QA Systems**, managed by Regional Offices (RO), which assess draft ToR and draft Reports.

UNICEF is looking for an institution to ensure the reviewing of and rating the quality of draft ToR, as well as draft and final evaluation reports supported by UNICEF country and regional offices all over the world, as well as HQ divisions.

Expected results

The selected institutions will review draft ToR, as well as draft and final evaluation reports in English, French and Spanish received by the EO and selected ROs (up to a maximum of 200 draft/final reports and 50 draft ToR in one year timeframe), rate them against UNEG/UNICEF standards, write an executive feedback to be sent to the CO concerned, and make analysis of trends, key weaknesses and strengths of UNICEF-managed/supported evaluation reports and ToRs.

Expected deliverables

Within the Global Evaluation Quality Oversight System, the selected Institution will deliver the following outputs:

A. Draft ToR and draft Reports (contract to be managed by Regional Offices)

A1: Draft Evaluation ToR and draft Evaluation reports reviewed, rated and executive feedback sent

UNICEF Country Offices are sending the draft ToR and reports to Regional Office for real-time quality review and practical comments on how to improve them. The institution will carry out such review in maximum 3 working days for the draft ToR and 5 working days for the draft reports. The institution will provide professional and practical feedback according to pre-agreed templates (see hyperlink below for the evaluation reports, and attachment for the ToR).

A2: Regional overview of evaluation draft ToR and draft reports reviewed

The institution will undertake an annual review of feedback provided (see attachment with the example of last year); identify lessons to be learned on evaluation ToR and reports. Will compare these results with those which emerged from the two previous yearly exercises undertaken in the region and will identify lessons to be learned, emerging good practices and actionable recommendations to improve the quality assurance system as well as the quality of Evaluations in the specific region assessed.

A3: Regional Evaluation Help Desk.

The objective is to ensure real-time trouble-shooting and ad hoc technical assistance to UNICEF Country Offices when requested, for instance providing a second review of ToR, specific technical notes, etc. Timing and content of any specific task to be agreed about beforehand with the RO concerned.

B. Final Reports (contract to be managed by EO)

B1: Final Evaluation reports reviewed, rated and executive feedback sent

Download the final reports from the UNICEF Intranet database, and review and rate final Evaluation reports received in English, French and Spanish against UNEG/UNICEF standards using the Feedback Template – both the comprehensive as well as the executive one (available at http://www.unicef.org/evaluation/files/Tool_2012_v2.xlsx), highlighting strengths, weaknesses and recommendations to improve the quality of future evaluations reports.

The estimated total number of final evaluation reports to be reviewed will be between a minimum of 50 and a maximum of 150 in one year timeframe, out of which about 80% in English, 15% in French and 5% in Spanish.

Reports must be fully rated and the feedback given within 10 working days of receipt. At times, there may be as many as 20 to be handled within the 10 day period. If reports to be rated within the 10 working days exceed 20, the rating time will be extended.

B2. Global analysis of trends, key weaknesses and strengths of reports reviewed

Every year, produce a Meta-evaluation based on the assessments of all final reports reviewed that year highlighting key trends, key weaknesses and strengths of reports reviewed, including lessons learned and good practices on Evaluation reports, and actionable conclusions and recommendations to improve GEREOS as well as the quality of Evaluation reports. Pls refer to the latest meta-evaluation for your easy reference at http://www.unicef.org/evaldatabase/index_60935.html. A PowerPoint highlighting key issues should also be prepared.

Management of the system

This Long Term Agreement covers a) the Global Evaluation Reports Oversight System (GEROS), managed by the Evaluation Office (EO), which rates final evaluation reports commissioned by UNICEF Country Offices (CO), Regional Offices and HQ divisions against the UNEG/UNICEF Evaluation Report Standards, and b) Regional QA Systems, managed by Regional Offices (RO), which assess draft ToR and draft Reports. However, the Evaluation Offices will raise a contract to cover the GEREOS system only, while Regional Offices (if any) will raise separate contracts to cover the Regional QA Systems.

The contract raised by the EO (covering the GEROS System) will be managed by the Senior Evaluation Specialist, Systemic strengthening, with the support of the Knowledge Management (KM) specialist. The contracts (if any) raised by Regional Offices will be managed by the respective Regional M&E Chiefs.

The selected institution will appoint a project manager who will ensure consistency of rating, quality and timely delivery of expected products, and overall coordination with UNICEF Evaluation Office. The project manager will also provide an update on a monthly basis, which will include a tracking matrix highlighting the status of reviews, ratings and executive feedback. The project manager will be the point of contact with UNICEF for any issues related to this Long Term Agreement.

Please note that, to avoid potential conflict of interest, the following will be applied:

- The company that will win this bid, will not review any ToR, draft and/or final evaluation reports of evaluation conducted by the same company
- The reviewer who rates the final evaluation reports will be different from the reviewer who rates the draft ToR and/or draft report

Qualifications

- Excellent and proved knowledge of evaluation methodologies and approaches
- Proven experience with Quality review of evaluation reports, preferably with UN agencies
- Proven practical professional experience in designing and conducting major evaluations
- Excellent analytical and writing skills in English required. Adequacy in French and Spanish required, with excellence in French and Spanish a strong advantage
- Familiarity with UNEG/UNICEF evaluation standards is an asset
- Sectorial knowledge in Child survival and development and at least other two UNICEF area of intervention (education; HIV/AIDS; Child protection; Social protection) in English language
- Knowledge and expertise of other or similar quality assurance systems will also be an asset
- Proven capacities in managing databases

Duration of contract

The Long Term Agreement will start 1 October 2012 and will expire 30 August 2015.

Appendix IV List of Reports Assessed

COUNTRY	REGION	YEAR	SEQ. #	TITLE	RATING	LANGUAGE
WCARO, Senegal	WCARO	2012	2012/008	Real-Time Independent Assessment (RTIA) of UNICEF's Response to the Sahel Food and Nutrition Crisis, 2011–2012	Outstanding, Best Practice	English
CEECIS Regional	CEE/CIS	2012	2012/001	Evaluation of the Roma Good Start Initiative	Outstanding, Best Practice	English
USA	HQ	2012	2012/006	IASC Real-Time Evaluation of the Humanitarian Response to the Horn of Africa Drought Crisis - Ethiopia	Highly Satisfactory	English
USA	HQ	2012	2012/008	IASC Real-Time Evaluation of the Humanitarian Response to the Horn of Africa Drought Crisis - Kenya	Highly Satisfactory	English
USA	HQ	2012	2012/007	IASC Real-Time Evaluation of the Humanitarian Response to the Horn of Africa Drought Crisis - Somalia	Highly Satisfactory	English
USA	HQ	2012	2012/005	IASC Real-Time Evaluation of the Humanitarian Response to the Horn of Africa Drought Crisis - Regional Mechanisms and Support during the Response	Highly Satisfactory	English
Democratic Republic of Congo	WCARO	2012	2012/002	Évaluation du programme 2007-2011 pour les Enfants Associés aux Forces et aux Groupes Armés en RDC	Highly Satisfactory	English
Nepal	ROSA	2012	2012/006	Evaluation of Conflict-Related Sexual Violence Project	Highly Satisfactory	English
USA	HQ	2012	2012/015	Review of the Fast Track Recruitment Process	Highly Satisfactory	English
Bosnia and Herzegovina	CEE/CIS	2012	2012/002	Joint Evaluation of Social Protection and Inclusion	Highly Satisfactory	English
Bosnia and Herzegovina	CEE/CIS	2012	2012/004	Final Evaluation of MDG-F Culture for Development Programme	Highly Satisfactory	English
Botswana	ESARO	2012	2012/004	Iplegeng Evaluation	Highly Satisfactory	English
Burkina Faso	WCARO	2012	2012/006	Évaluation finale du projet « Approvisionnement en Eau Potable, Assainissement de base et Hygiène dans les provinces du Ganzourgou et de la Gnagna, Burkina Faso – 2007-2010 »	Highly Satisfactory	French
Cambodia	EAPRO	2012	2012/007	Buddhist Leadership Initiative Evaluation 2008-2012	Highly Satisfactory	English
Chad	WCARO	2012	2012/004	Évaluation de la prise en charge communautaire de la malnutrition aiguë (PCMA) : étude de cas du Tchad	Highly Satisfactory	French
Democratic Republic of Congo	WCARO	2012	2012/004	Évaluation du Programme Éducation de base 2008-2012	Highly Satisfactory	French

COUNTRY	REGION	YEAR	SEQ. #	TITLE	RATING	LANGUAGE
India	ROSA	2012	2012/011	Evaluation of Knowledge Community on Children in India Internship Programme	Highly Satisfactory	English
India	ROSA	2012	2012/015	Evaluation of District Planning and Monitoring Unit	Highly Satisfactory	English
Lebanon	MENARO	2012	2012/003	Evaluation of Child Friendly Community Initiative	Highly Satisfactory	English
Liberia	WCARO	2012	2012/002	Evaluation of MoHSW & Bomi CHT Performance based contracting	Highly Satisfactory	English
Macedonia	CEE/CIS	2012	2012/007	Strengthening National Capacities to Prevent Domestic Violence	Highly Satisfactory	English
Macedonia	CEE/CIS	2012	2012/008	Enhancing Inter-Ethnic Community Dialogue and Collaboration	Highly Satisfactory	English
Moldova	CEE/CIS	2012	2012/001	Final Evaluation of the "Reform of Juvenile Justice System in Moldova" Project	Highly Satisfactory	English
Moldova	CEE/CIS	2012	2012/002	External Evaluation of the "Child-Friendly School" Initiative (2007-2011) in the Republic of Moldova	Highly Satisfactory	English
Nepal	ROSA	2012	2012/001	Evaluation of Community Management of Acute Malnutrition Community (CMAM): Nepal Case Studies	Highly Satisfactory	English
Occupied Palestinian Territ.	MENARO	2012	2012/001	Evaluation of the Child-Friendly Schools	Highly Satisfactory	English
Senegal	WCARO	2012	2012/007	Évaluation des systèmes statistiques de routine de la région de Kolda	Highly Satisfactory	French
Sierra Leone	WCARO	2012	2012/001	Evaluation of UNICEF role as a Lead Partner in Education	Highly Satisfactory	English
Somalia	ESARO	2012	2012/011	Final Evaluation of UNICEF's Programme "Making PPP Work for Rural Water Supply in Somalia"	Highly Satisfactory	English
Somalia	ESARO	2012	2012/001	Developing a Local Model for The Delivery of Primary Education in Karkaar Region, Puntland	Highly Satisfactory	English
Uganda	ESARO	2012	2012/007	WASH Initiative for the Rural Poor in 21 Districts in Uganda: End of Term Evaluation	Highly Satisfactory	English
United Rep. of Tanzania	ESARO	2012	2011/004	Evaluation of Government of Tanzania/UNICEF 7 Learning Districts Strategy (2007-2011)	Highly Satisfactory	English
USA	HQ	2012	2012/011	Global Evaluation of Life Skills Education Programmes	Highly Satisfactory	English
USA	HQ	2012	2012/010	Protecting Children from Violence: A Synthesis of Evaluation Findings	Highly Satisfactory	English

COUNTRY	REGION	YEAR	SEQ. #	TITLE	RATING	LANGUAGE
Zambia	ESARO	2012	2012/011	Evaluation of the UN Joint Programme on Human Trafficking.	Highly Satisfactory	English
Zimbabwe	ESARO	2012	2012/002	Evaluation of the Use of Point of Care PIMA CD4 Cell Count Machines for HIV Positive Women and their Families in Maternal, Newborn and Child Health (MNCH) Settings in Seven Districts in Zimbabwe	Highly Satisfactory	English
Honduras	TACRO	2012	2012/001	Municipal Programs on Children, Adolescents and Youth. PMIAJ COMVIDA in Honduras	Highly Satisfactory	Spanish
Ghana	WCARO	2012	2012/043	Evaluation of the Government of Ghana - UNICEF Integrated Approach to Guinea Worm eradication through Water Supply, Sanitation and Hygiene in Northern Region, Ghana (I-WASH)	Highly Satisfactory	English
Bulgaria	CEE/CIS	2012	2012/002	Evaluation of the project "Expansion of foster care model in Bulgaria"	Highly Satisfactory	English
Sudan	MENARO	2012	2012/001	Programme Evaluation	Highly Satisfactory	English
Benin	WCARO	2012	2012/003	Évaluation finale du projet de formation diplômante des enseignants ex-communautaires reverses contractuels de l'état	Highly Satisfactory	French
Croatia	CEE/CIS	2012	2012/001	External evaluation of the "For Safe and Enabling School Environment"	Highly Satisfactory	English
Pakistan	ROSA	2012	2012/008	Evaluation of Community Management of Acute Malnutrition (CMAM): Pakistan Country Case Study.	Highly Satisfactory	English
ESARO , Kenya	ESARO	2012	2012/001	Building a culture of resilience	Highly Satisfactory	English
Moldova	CEE/CIS	2012	2012/003	Evaluation of Implementation of the National Strategy & Action Plan for the Reform of the Residential Childcare System in Moldova 2007-2012	Highly Satisfactory	English
Ethiopia	ESARO	2012	2011/049	Evaluation of Adolescent Development Programme	Highly Satisfactory	English
USA (managed by Universalialia)	HQ	2012	2012/003	Global Evaluation of the Application of the Human Rights-Based Approach to UNICEF Programming	Highly Satisfactory	English
USA (managed by Universalialia)	HQ	2013	2013/001	The Global Evaluation of Emergency Response Funds (ERFs)	Highly Satisfactory	English
Democratic Republic of Congo	WCARO	2012	2012/001	Évaluation externe du Programme Élargi d'appui aux retours (PEAR+) à l'Est de la RDC	Mostly Satisfactory	French
USA	HQ	2012	2012/004	Synthesis report of the IA RTEs of the Response to the Horn of Africa Drought	Mostly Satisfactory	English

COUNTRY	REGION	YEAR	SEQ. #	TITLE	RATING	LANGUAGE
Bosnia and Herzegovina	CEE/CIS	2012	2012/001	Evaluation of the Juvenile Justice	Mostly Satisfactory	English
China	EAPRO	2012	2012/008	Report of Final Evaluation of (MDGF) Joint Programme: UN-China Protecting and Promoting the Rights of China's Vulnerable Young Migrants (YEM)	Mostly Satisfactory	English
Cote D'Ivoire	WCARO	2012	2011/010	Evaluation of the measles campaign's best practices	Mostly Satisfactory	French
Democratic Republic of Congo	WCARO	2012	2012/003	Évaluation du programme École et Village Assainis	Mostly Satisfactory	French
Ethiopia	ESARO	2012	2012/001	End Term Evaluation of Capacity Building Strategy in remote zones in SNNPR with teams of NUNV experts	Mostly Satisfactory	English
Iran	MENARO	2012	2012/002	The Adolescent Friendly Services (AFS) project evaluation	Mostly Satisfactory	English
Lesotho	ESARO	2012	2012/001	Support to Lesotho HIV and AIDS Response: Empowerment of Orphans and other Vulnerable Children, Final Evaluation	Mostly Satisfactory	English
Liberia	WCARO	2012	2012/003	Evaluation of WASH interventions in Urban Slums of Monrovia and Buchanan 2011-2012	Mostly Satisfactory	English
Liberia	WCARO	2012	2012/004	Evaluation of the Accelerated Learning Program in Liberia	Mostly Satisfactory	English
Republic of Mozambique	ESARO	2012	2012/006	End Cycle Evaluation of the Child Friendly Schools Programme, 2006-2011	Mostly Satisfactory	English
Sierra Leone	WCARO	2012	2012/002	Mid Term Evaluation of the National Gender Strategic Plan	Mostly Satisfactory	English
Somalia	ESARO	2012	2012/010	TOSTAN Pilot Project on "Ending FGM/C" in Northwest and Northeast Zone in Somalia	Mostly Satisfactory	English
South Africa	ESARO	2012	2012/001	The South African Child Support Grant Impact Assessment	Mostly Satisfactory	English
Zimbabwe	ESARO	2012	2012/001	Evaluation of the Basic Education Assistance Module Programme	Mostly Satisfactory	English
Barbados	TACRO	2012	2012/001	Evaluation of the impact of the Child Friendly Schools approach in the Eastern Caribbean	Mostly Satisfactory	English
Costa Rica	TACRO	2012	2012/001	Final Evaluation of the Joint Programme (UNICEF, FAO, UNFPA, UNESCO, and IOM): "Youth, Employment and Migration (JEM): The Single-Window turn-key solution for Youth Employment in Desamparados and Upala".	Mostly Satisfactory	Spanish
Gambia	WCARO	2012	2012/001	Evaluation of the Girls Education Project of the Forum for African Women Educationalists - The Gambia (FAWEGAM)	Mostly Satisfactory	English

COUNTRY	REGION	YEAR	SEQ. #	TITLE	RATING	LANGUAGE
Republic of Cameroon	WCARO	2012	2012/003	Évaluation du programme de prise en charge de la malnutrition	Mostly Satisfactory	French
Kenya	ESARO	2012	2012/004	Evaluation of Integrated Management of Acute Malnutrition (IMAM): Kenya Country Case Study	Mostly Satisfactory	English
Ethiopia	ESARO	2012	2011/047	Final Evaluation of the Social Cash Transfer Scheme of UNICEF	Mostly Satisfactory	English
Ethiopia	ESARO	2012	2011/058	Evaluation of Community Management of Acute Malnutrition (CMAM)	Mostly Satisfactory	English
Pakistan	ROSA	2012	2012/004	Evaluation of Social Reintegration of Street Children Project	Mostly Satisfactory	English
Ukraine	CEE/CIS	2012	2012/002	'Prevention of Mother-to-Child Transmission and Improving Neonatal Outcomes among Drug-Dependent Pregnant Women and Children Born to Them in Three Cities in Ukraine'	Mostly Satisfactory	English
Liberia	WCARO	2012	2012/010	MULTI-SECTOR EMERGENCY RESPONSE TO IVORIAN REFUGEES AND HOST COMMUNITIES IN FOUR LIBERIAN COUNTIES OF NIMBA, GRAND GEDEH, RIVERGEE & MARYLAND	Unsatisfactory	English
Bolivia	TACRO	2012	2012/009	Bolivian Health Sector Support Program – PASS, End-of-Project Evaluation, Final Report.	Unsatisfactory	English
Bosnia and Herzegovina	CEE/CIS	2012	2012/003	Youth Employability and Retention Programme	Unsatisfactory	English
East Timor	EAPRO	2012	2012/002	JOINT PROGRAMME PROMOTING SUSTAINABLE FOOD AND NUTRITION SECURITY IN TIMOR-LESTE FINAL EVALUATION	Unsatisfactory	English
Zambia	ESARO	2012	2012/012	Summative Evaluation of the International Inspiration Project on HIV and AIDS and Life-skills through Sports.	Unsatisfactory	English
Ecuador	TACRO	2012	2012/001	Child Labor in Ecuador: an evaluation	Unsatisfactory	Spanish

Appendix V Review tool

UNICEF Global Evaluation Report Oversight System (GEROS) Review Template

Colour Coding	CC	Dark green	Green	Amber	Red	White
	Questions	Outstanding	Yes	Mostly Satisfactory	No	Not Applicable
	Section & Overall Rating	Outstanding, best practice	Highly Satisfactory	Mostly Satisfactory	Unsatisfactory	

The key questions are highlighted as shown here, and are important questions in guiding the analysis of the section

The Cornerstone questions are in column J and are questions that need to be answered for rating and justification of each of the six sections

[UNEG Standards for Evaluation in the UN System](#)
 [UNEG Norms for Evaluation in the UN System](#)
 [UNICEF Adapted UNEG Evaluation Report Standards](#)

	Response				
Title of the Evaluation Report					
Report sequence number		Date of Review		Year of the Evaluation Report	
Region				Country	
Type of Report				ToRs Present	
Name of reviewer					
Classification of Evaluation Report					Comments
Geographical <i>(Coverage of the programme being evaluated & generalizability of evaluation findings)</i>					
Management of Evaluation <i>(Managerial control and oversight of evaluation decisions)</i>					

Purpose <i>(Speaks to the overarching goal for conducting the evaluation; its raison d'être)</i>				
Result <i>(Level of changes sought, as defined in RBM: refer to substantial use of highest level reached)</i>				
MTSP Correspondence <i>(Alignment with MTSP focus area priorities: (1) Young child survival and development; (2) Basic education and gender equality; (3) HIV/AIDS and children; (4) Child protection from violence, exploitation and abuse; and (5) Policy advocacy and partnerships for children's rights)</i>				
Level of Independence <i>(Implementation and control of the evaluation activities)</i>				
Stage				
SECTION A: OBJECT OF THE EVALUATION				
Question	cc	Remarks	A/ Does the report present a clear & full description of the 'object' of the evaluation? The report should describe the object of the evaluation including the results chain, meaning the 'theory of change' that underlies the programme being evaluated. This theory of	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>
Object and context				
1 Is the object of the evaluation well described? This needs to include a clear description of the interventions (project, programme, policies, otherwise) to be evaluated including how the designer thought that it would address the problem identified, implementing modalities, other parameters including costs, relative importance in the organization and (number of) people reached.				

			<p>change includes what the programme was meant to achieve and the pathway (chain of results) through which it was expected to achieve this.</p> <p>The context of key social, political, economic, demographic, and institutional factors that have a direct bearing on the object should be described. For example, the partner government's strategies and priorities, international, regional or country development goals, strategies and frameworks, the concerned agency's corporate goals & priorities, as appropriate.</p>	
<p>2 Is the context explained and related to the object that is to be evaluated?</p> <p>The context includes factors that have a direct bearing on the object of the evaluation: social, political, economic, demographic, institutional. These factors may include strategies, policies, goals, frameworks & priorities at the: international level; national Government level; individual agency level</p>				
<p>3 Does this illuminate findings?</p> <p>The context should ideally be linked to the findings so that it is clear how the wider situation may have influenced the outcomes observed.</p>				
Theory of Change				
<p>4 Is the results chain or logic well articulated?</p> <p>The report should identify how the designers of the evaluated object thought that it would address the problem that they had identified. This can include a results chain or other logic models such as theory of change. It can include inputs, outputs and outcomes, it may also include impacts. The models need to be clearly described and explained.</p>				
Stakeholders and their contributions				
<p>5 Are key stakeholders clearly identified?</p> <p>These include o implementing agency(ies) o development partners o rights holders o primary duty bearers o secondary duty bearers</p>				
<p>6 Are key stakeholders' contributions described?</p> <p>This can involve financial or other contributions and should be specific. If joint program also specify UNICEF contribution, but if basket funding question is not applicable</p>				
<p>7 Are UNICEF contributions described?</p> <p>This can involve financial or other contributions and should be specific</p>				

Implementation Status					
8 Is the implementation status described? This includes the phase of implementation and significant changes that have happened to plans, strategies, performance frameworks, etc that have occurred - including the implications of these changes					
Executive Feedback on Section A Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>					
SECTION B: EVALUATION PURPOSE, OBJECTIVES AND SCOPE					
Question	cc	Remarks	B/ Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation?	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>	
Purpose, objectives and scope			The purpose of the evaluation should be clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed, and how the information will be used.		
9 Is the purpose of the evaluation clear? This includes why the evaluation is needed at this time, who needs the information, what information is needed, how the information will be used.			The report should provide a clear explanation of the evaluation objectives and scope including main evaluation questions and describes and justifies what the evaluation did and did not cover. The report should describe and provide an explanation of the chosen evaluation criteria, performance standards, or other criteria used by the evaluators.		
10 Are the objectives and scope of the evaluation clear and realistic? This includes: Objectives should be clear and explain what the evaluation is seeking to achieve; Scope should clearly describe and justify what the evaluation will and will not cover; Evaluation questions may optionally be included to add additional details					
11 Do the objective and scope relate to the purpose? The reasons for holding the evaluation at this time in the project cycle (purpose) should link logically with the specific objectives the evaluation seeks to achieve and the boundaries chosen for the evaluation (scope)					

Evaluation framework			
<p>12 Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the Purpose?</p> <p>It is imperative to make the basis of the value judgements used in the evaluation transparent if it is to be understood and convincing. UNEG evaluation standards refer to the OECD/DAC criteria, but other criteria can be used such as Human rights and humanitarian criteria and standards (e.g. SPHERE Standards) but this needs justification.. Not all OECD/DAC criteria are relevant to all evaluation objectives and scopes. The ToRs may set the criteria to be used, but these should be (re)confirmed by the evaluator. Standard OECD DAC Criteria include: Relevance; Effectiveness; Efficiency; Sustainability; Impact Additional humanitarian criteria include; Coverage; Coordination; Coherence; Protection; timeliness; connectedness; appropriateness. <i>(This is an extremely important question to UNICEF)</i></p>			
<p>13 Does the evaluation explain why the evaluation criteria were chosen and/or any standard DAC evaluation criteria (above) rejected?</p> <p>The rationale for using each particular criterion and rejecting any standard OECD-DAC criteria (where they would be applicable) should be explained in the report.</p>			
<p>Executive Feedback on Section B</p> <p>Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i></p>			

SECTION C: EVALUATION METHODOLOGY, GENDER, HUMAN RIGHTS AND EQUITY						
Question	cc	Remarks	<p>C/ Is the methodology appropriate and sound?</p> <p>The report should present a transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve the evaluation purposes.</p> <p>The report should also present a sufficiently detailed description of methodology in which methodological choices are made explicit and justified and in which limitations of methodology applied are included. The report should give the elements to assess the appropriateness of the methodology. Methods as such are not 'good' or 'bad', they are only so in relation to what one tries to get to know as part of an evaluation. Thus this standard assesses the suitability of the methods selected for the specifics of the evaluation concerned, assessing if the methodology is suitable to the subject matter and the information collected are sufficient to meet the evaluation objectives.</p>	<p>Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i></p>		
Data collection						
<p>14 Does the report specify data collection methods, analysis methods, sampling methods and benchmarks? This should include the rationale for selecting methods and their limitations based on commonly accepted best practice.</p>						
<p>15 Does the report specify data sources, the rationale for their selection, and their limitations? This should include a discussion of how the mix of data sources was used to obtain a diversity of perspectives, ensure accuracy & overcome data limits</p>						
Ethics						
<p>16 Are ethical issues and considerations described? The design of the evaluation should contemplate: How ethical the initial design of the programme was; The balance of costs and benefits to participants (including possible negative impact) in the programme and in the evaluation; The ethics of who is included and excluded in the evaluation and how this is done</p>						
<p>17 Does the report refer to ethical safeguards appropriate for the issues described? When the topic of an evaluation is contentious, there is a heightened need to protect those participating. These should be guided by the UNICEF Evaluation Office Technical Note and include: protection of confidentiality; protection of rights; protection of dignity and welfare of people (especially children); Informed consent; Feedback to participants; Mechanisms for shaping the behaviour of evaluators and data collectors</p>						

Results Based Management				
18 Is the capability and robustness of the evaluated object's monitoring system adequately assessed? The evaluation should consider the details and overall functioning of the management system in relation to results: from the M&E system design, through individual tools, to the use of data in management decision making.				
19 Does the evaluation make appropriate use of the M&E framework of the evaluated object? In addition to articulating the logic model (results chain) used by the programme, the evaluation should make use of the object's logframe or other results framework to guide the assessment. The results framework indicates how the programme design team expected to assess effectiveness, and it forms the guiding structure for the management of implementation.				
Human Rights, Gender and Equity				
20 Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity? This could be done in a variety of ways including: use of a rights-based framework, use of CRC, CCC, CEDAW and other rights related benchmarks, analysis of right holders and duty bearers and focus on aspects of equity, social exclusion and gender. Style includes: using human-rights language; gender-sensitive and child-sensitive writing; disaggregating data by gender, age and disability groups; disaggregating data by socially excluded groups				
21 Does the evaluation assess the extent to which the implementation of the evaluated object was monitored through human rights (inc. gender, equity & child rights) frameworks? UNICEF commits to go beyond monitoring the achievement of desirable outcomes, and to ensure that these are achieved through morally acceptable processes. The evaluation should consider whether the programme was managed and adjusted according to human rights and gender monitoring of processes.				

<p>22 Do the methodology, analytical framework, findings, conclusions, recommendations & lessons provide appropriate information on HUMAN RIGHTS (inc. women & child rights)? The inclusion of human rights frameworks in the evaluation methodology should continue to cascade down the evaluation report and be obvious in the data analysis, findings, conclusions, any recommendations and any lessons learned. If identified in the scope the methodology should be capable of assessing the level of: Identification of the human rights claims of rights-holders and the corresponding human rights obligations of duty-bearers, as well as the immediate underlying & structural causes of the non realisation of rights.; Capacity development of rights-holders to claim rights, and duty-bearers to fulfil obligations.</p>					
<p>23 Do the methodology, analytical framework, findings, conclusions, recommendations & lessons provide appropriate information on GENDER EQUALITY AND WOMEN'S EMPOWERMENT? The inclusion of gender equality frameworks in the evaluation methodology should continue to cascade down the evaluation report and be obvious in the data analysis, findings, conclusions, any recommendations and any lessons learned. If identified in the scope the methodology should be capable of assessing the immediate underlying & structural causes of social exclusion; and capacity development of women to claim rights, and duty-bearers to fulfil their equality obligations.</p>					
<p>24 Do the methodology, analytical framework, findings, conclusions, recommendations & lessons provide appropriate information on EQUITY? The inclusion of equity considerations in the evaluation methodology should continue to cascade down the evaluation report and be obvious in the data analysis, findings, conclusions, any recommendations and any lessons learned. If identified in the scope the methodology should be capable of assessing the capacity development of rights-holders to claim rights, and duty-bearers to fulfil obligations & aspects of equity.</p>					

Stakeholder participation				
25 Are the levels and activities of stakeholder consultation described? This goes beyond just using stakeholders as sources of information and includes the degree of participation in the evaluation itself. The report should include the rationale for selecting this level of participation. Roles for participation might include: o Liaison o Technical advisory o Observer o Active decision making The reviewer should look for the soundness of the description and rationale for the degree of participation rather than the level of participation itself.				
26 Are the levels of participation appropriate for the task in hand? The breadth & degree of stakeholder participation feasible in evaluation activities will depend partly on the kind of participation achieved in the evaluated object. The reviewer should note here whether a higher degree of participation may have been feasible & preferable.				
Methodological robustness				
27 Is there an attempt to construct a counterfactual or address issues of contribution/attribution? The counterfactual can be constructed in several ways which can be more or less rigorous. It can be done by contacting eligible beneficiaries that were not reached by the programme, or a theoretical counterfactual based on historical trends, or it can also be a comparison group.				
28 Does the methodology facilitate answers to the evaluation questions in the context of the evaluation? The methodology should link back to the Purpose and be capable of providing answers to the evaluation questions.				
29 Are methodological limitations acceptable for the task in hand? Limitations must be specifically recognised and appropriate efforts taken to control bias. This includes the use of triangulation, and the use of robust data collection tools (interview protocols, observation tools etc). Bias limitations can be addressed in three main areas: Bias inherent in the sources of data; Bias introduced through the methods of data collection; Bias that colours the interpretation of findings				

Executive Feedback on Section C Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>				
SECTION D: FINDINGS AND CONCLUSIONS				
Question	cc	Remarks	D/ Are the findings and conclusions, clearly presented, relevant and based on evidence & sound analysis?	
Completeness and logic of findings				
30 Are findings clearly presented and based on the objective use of the reported evidence? Findings regarding the inputs for the completion of activities or process achievements should be distinguished clearly from results. Findings on results should clearly distinguish outputs, outcomes and impacts (where appropriate). Findings must demonstrate full marshalling and objective use of the evidence generated by the evaluation data collection. Findings should also tell the 'whole story' of the evidence and avoid bias.			Findings should respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report. They should be based on evidence derived from data collection and analysis methods described in the methodology section of the report. Conclusions should present reasonable judgments based on findings and substantiated by evidence, providing insights pertinent to the object and purpose of the evaluation.	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>
31 Do the findings address all of the evaluation's stated criteria and questions? The findings should seek to systematically address all of the evaluation questions according to the evaluation framework articulated in the report.				
32 Do findings demonstrate the progression to results based on the evidence reported? There should be a logical chain developed by the findings, which shows the progression (or lack of) from implementation to results.				

<p>33 Are gaps and limitations discussed? The data may be inadequate to answer all the evaluation questions as satisfactorily as intended, in this case the limitations should be clearly presented and discussed. Caveats should be included to guide the reader on how to interpret the findings. Any gaps in the programme or unintended effects should also be addressed.</p>					
<p>34 Are unexpected findings discussed? If the data reveals (or suggests) unusual or unexpected issues, these should be highlighted and discussed in terms of their implications.</p>					
Cost Analysis					
<p>35 Is a cost analysis presented that is well grounded in the findings reported? Cost analysis is not always feasible or appropriate. If this is the case then the reasons should be explained. Otherwise the evaluation should use an appropriate scope and methodology of cost analysis to answer the following questions: o How programme costs compare to other similar programmes or standards o Most efficient way to get expected results o Cost implications of scaling up or down o Cost implications for replicating in a different context o Is the programme worth doing from a cost perspective o Costs and the sustainability of the programme.</p>					
Contribution and causality					
<p>36 Does the evaluation make a fair and reasonable attempt to assign contribution for results to identified stakeholders? For results attributed to the programme, the result should be mapped as accurately as possible to the inputs of different stakeholders.</p>					
<p>37 Are causal reasons for accomplishments and failures identified as much as possible? These should be concise and usable. They should be based on the evidence and be theoretically robust. <i>(This is an extremely important question to UNICEF)</i></p>					

Strengths, weaknesses and implications				
38 Are the future implications of continuing constraints discussed? The implications can be, for example, in terms of the cost of the programme, ability to deliver results, reputational risk, and breach of human rights obligations.				
39 Do the conclusions present both the strengths and weaknesses of the evaluated object? Conclusions should give a balanced view of both the stronger aspects and weaker aspects of the evaluated object with reference to the evaluation criteria and human rights based approach.				
Completeness and insight of conclusions				
40 Do the conclusions represent actual insights into important issues that add value to the findings? Conclusions should go beyond findings and identify important underlying problems and/or priority issues. Simple conclusions that are already well known do not add value and should be avoided.				
41 Do conclusions take due account of the views of a diverse cross-section of stakeholders? As well as being logically derived from findings, conclusions should seek to represent the range of views encountered in the evaluation, and not simply reflect the bias of the individual evaluator. Carrying these diverse views through to the presentation of conclusions (considered here) is only possible if the methodology has gathered and analysed information from a broad range of stakeholders.				
42 Are the conclusions pitched at a level that is relevant to the end users of the evaluation? Conclusions should speak to the evaluation participants, stakeholders and users. These may cover a wide range of groups and conclusions should thus be stated clearly and accessibly: adding value and understanding to the report (for example, some stakeholders may not understand the methodology or findings, but the conclusions should clarify what these findings mean to them in the context of the programme).				

Executive Feedback on Section D Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>				
SECTION E: RECOMMENDATIONS AND LESSONS LEARNED				
Question	cc	Remarks	E/ Are the recommendations and lessons learned relevant and actionable?	Constructive feedback for future reports <i>Including how to address weaknesses and maintaining good practice</i>
Relevance and clarity of recommendations				
43 Are the recommendations well-grounded in the evidence and conclusions reported? Recommendations should be logically based in findings and conclusions of the report.			Recommendations should be relevant and actionable to the object and purpose of the evaluation, be supported by evidence and conclusions, and be developed with involvement of relevant stakeholders. Recommendations should clearly identify the target group for each recommendation, be clearly stated with priorities for action, be actionable and reflect an understanding of the commissioning organization and potential constraints to follow up.	
44 Are recommendations relevant to the object and the purpose of the evaluation? Recommendations should be relevant to the evaluated object				
45 Are recommendations clearly stated and prioritised? If the recommendations are few in number (up to 5) then this can also be considered to be prioritised. Recommendations that are over-specific or represent a long list of items are not of as much value to managers. Where there is a long list of recommendations, the most important should be ordered in priority.				
Usefulness of recommendations				
46 Does each recommendation clearly identify the target group for action? Recommendations should provide clear and relevant suggestions for action linked to the stakeholders who might put that recommendation into action. This ensures that the evaluators have a good understanding of the programme dynamics and that recommendations are realistic.				

<p>47 Are the recommendations realistic in the context of the evaluation? This includes: o an understanding of the commissioning organisation o awareness of the implementation constraints o an understanding of the follow-up processes</p>				
<p>48 Does the report describe the process followed in developing the recommendations? The preparation of recommendations needs to suit the evaluation process. Participation by stakeholders in the development of recommendations is strongly encouraged to increase ownership and utility.</p>				
<p>Appropriate lessons learned</p>				
<p>49 Are lessons learned correctly identified? Lessons learned are contributions to general knowledge. They may refine or add to commonly accepted understanding, but should not be merely a repetition of common knowledge. Findings and conclusions specific to the evaluated object are not lessons learned.</p>				
<p>50 Are lessons learned generalised to indicate what wider relevance they may have? Correctly identified lessons learned should include an analysis of how they can be applied to contexts and situations outside of the evaluated object.</p>				
<p>Executive Feedback on Section E Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i></p>				

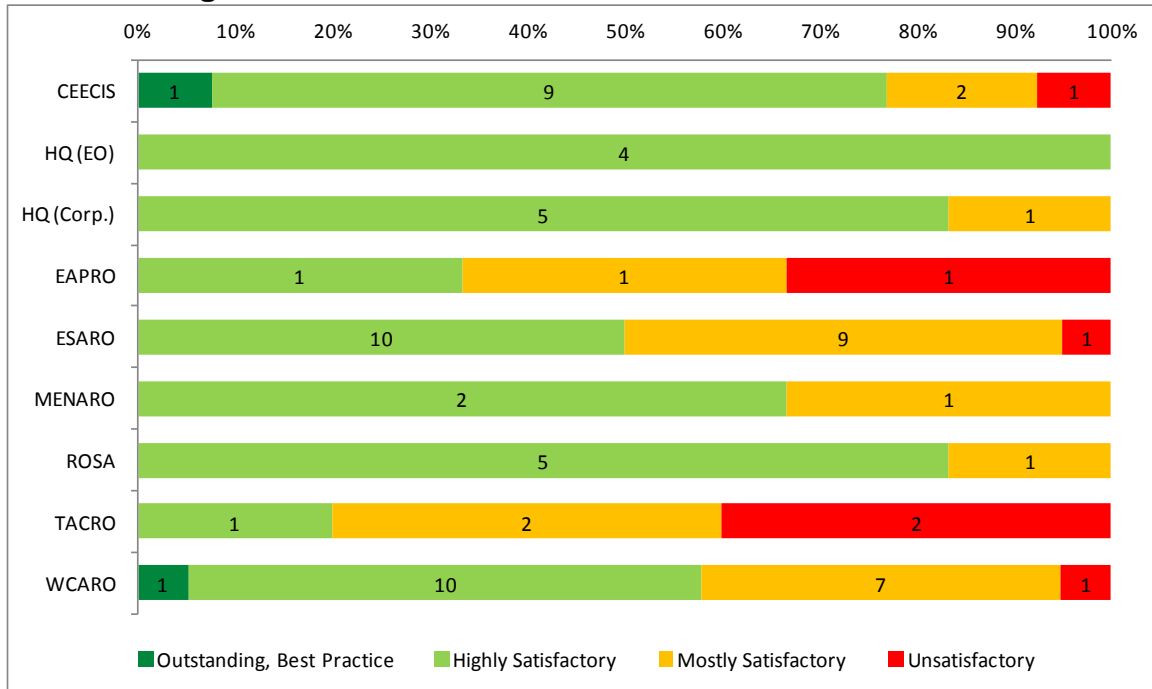
SECTION F: REPORT IS WELL STRUCTURED, LOGIC AND CLEAR				
Question	cc	Remarks		
Style and presentation				
51. Do the opening pages contain all the basic elements? Basic elements include all of: Name of the evaluated object; Timeframe of the evaluation and date of the report; Locations of the evaluated object; Names and/or organisations of evaluators; Name of the organisation commissioning the evaluation; Table of contents including tables, graphs, figures and annex; List of acronyms			F/ Overall, do all these elements come together in a well structured, logical, clear and complete report? The report should be logically structured with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations). It should read well and be focused.	
52 Is the report logically structured? Context, purpose, methodology and findings logically structured. Findings would normally come before conclusions, recommendations & lessons learnt				
53 Do the annexes contain appropriate elements? Appropriate elements may include: ToRs; List of interviewees and site visits; List of documentary evidence; Details on methodology; Data collection instruments; Information about the evaluators; Copy of the evaluation matrix; Copy of the Results chain. Where they add value to the report				
54 Do the annexes increase the usefulness and credibility of the report?				
Executive Summary				
55. Is an executive summary included as part of the report? If the answer is No, question 56 to 58 should be N/A				
56 Does the executive summary contain all the necessary elements? Necessary elements include all of: Overview of the evaluated object; Evaluation objectives and intended audience; Evaluation methodology; Most important findings and conclusions; Main recommendations				

<p>57 Can the executive summary stand alone? It should not require reference to the rest of the report documents and should not introduce new information or arguments</p>				
<p>58 Can the executive summary inform decision making? It should be short (ideally 2-3 pages), and increase the utility for decision makers by highlight key priorities.</p>				
<p>Executive Feedback on Section F Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i></p>				
<p>Additional Information</p>				
<p>Question</p>	<p>Remarks</p>			
<p>i/ Does the evaluation successfully address the Terms of Reference? If the report does not include a ToRs then a recommendation should be given to ensure that all evaluations include the ToRs in the future. Some evaluations may be flawed because the ToRs are inappropriate, too little time etc. Or, they may succeed despite inadequate ToRs. This should be noted under vii in the next section</p>				
<p>ii/ Identify aspects of good practice in the evaluation In terms of evaluation</p>				
<p>iii/ Identify aspects of good practice of the evaluation In terms of programmatic, sector specific, thematic expertise</p>				

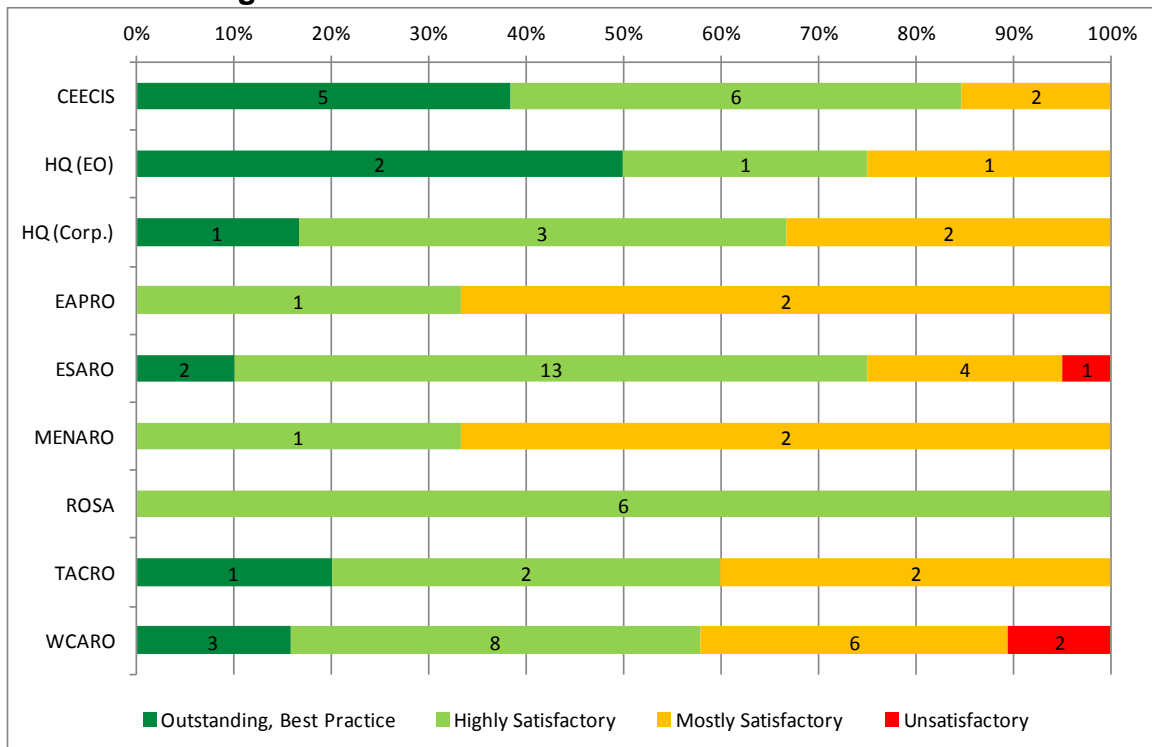
OVERALL RATING			
Question	cc	Remarks	OVERALL RATING Informed by the answers above, apply the reasonable person test to answer the following question: Ω/ Is this a credible report that addresses the evaluation purpose and objectives based on evidence, and that can therefore be used with confidence? This question should be considered from the perspective of UNICEF strategic management.
i/ To what extent does each of the six sections of the evaluation provide sufficient credibility to give the reasonable person confidence to act? Taken on their own, could a reasonable person have confidence in each of the five core evaluation elements separately? It is particularly important to consider: <ul style="list-style-type: none"> o Is the report methodologically appropriate? o Is the evidence sufficient, robust and authoritative? o Do the analysis, findings, conclusions and recommendations hold together? 			
ii/ To what extent do the six sections hold together in a logically consistent way that provides common threads throughout the report? The report should hold together not just as individually appropriately elements, but as a consistent and logical 'whole'.			
iii/ Are there any reasons of note that might explain the overall performance or particular aspects of this evaluation report? This is a chance to note mitigating factors and/or crucial issues apparent in the review of the report.	ToRs		
	Other		
Executive Feedback on Overall Rating Issues for this section relevant for feedback to senior management (positives & negatives), & justify rating. <i>Up to two sentences</i>			

Appendix VI Regional Ratings Graphs

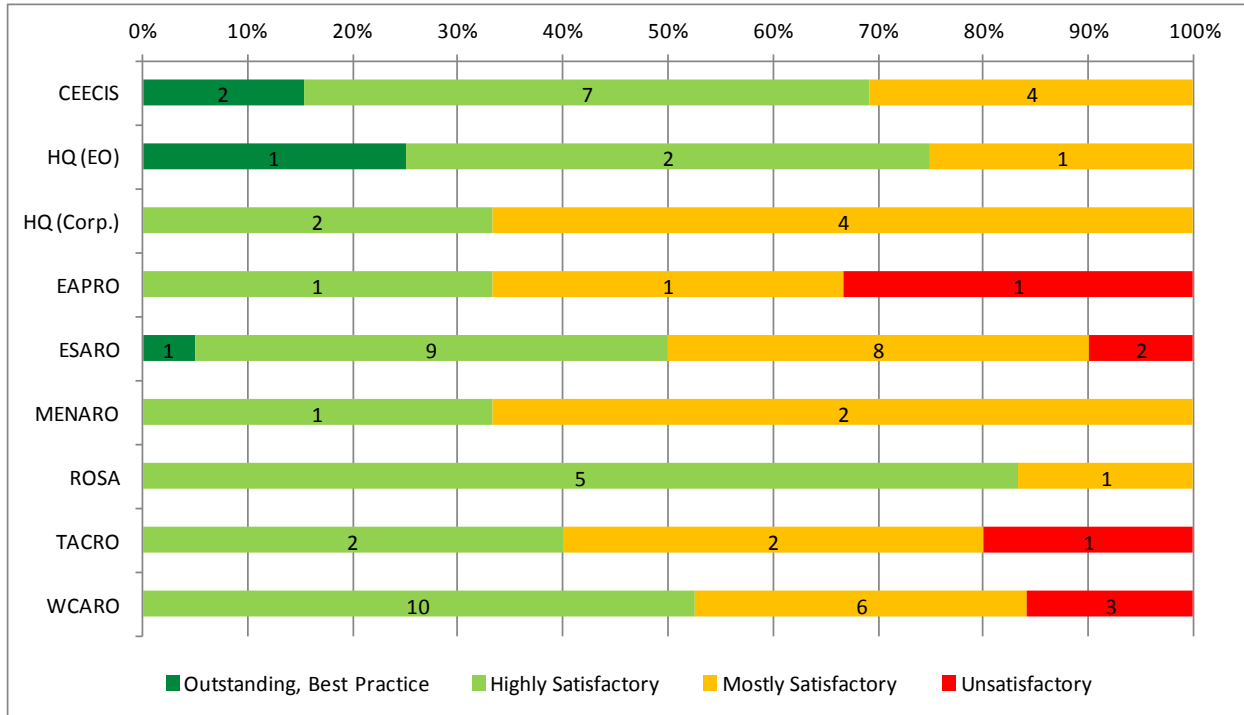
Overall Ratings



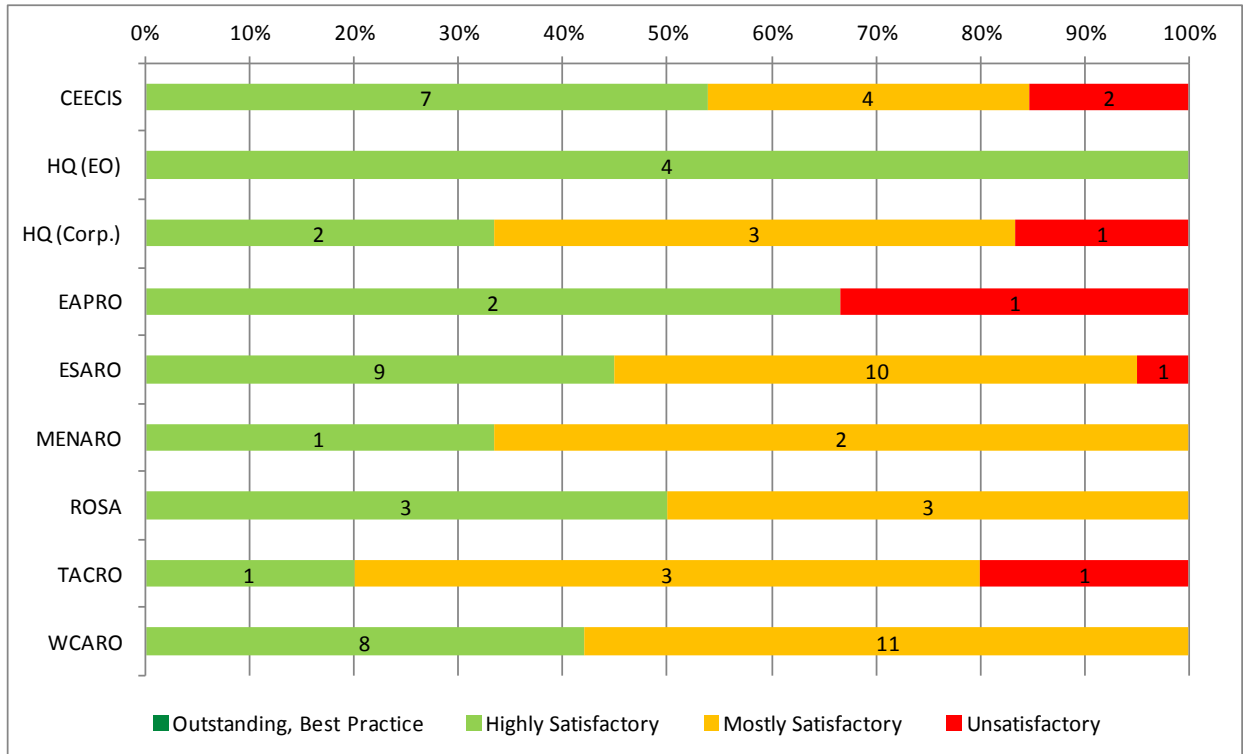
Section A Ratings



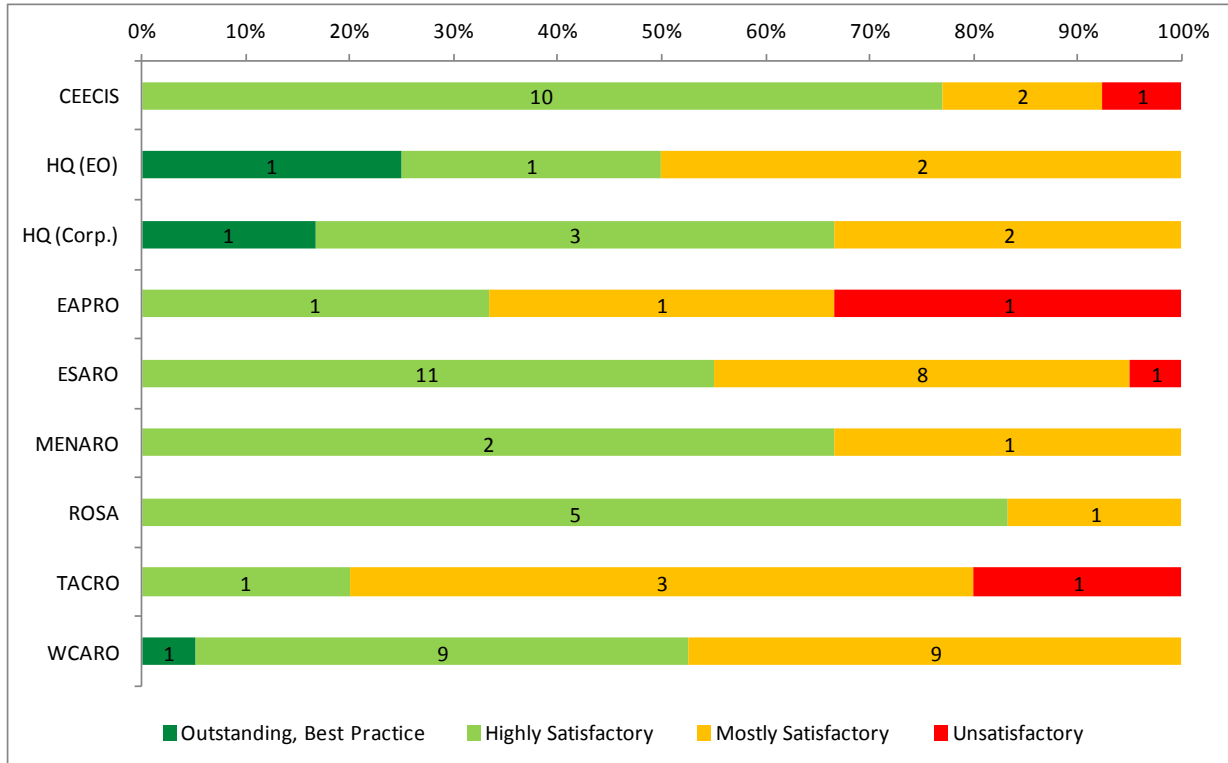
Section B Ratings



Section C Ratings



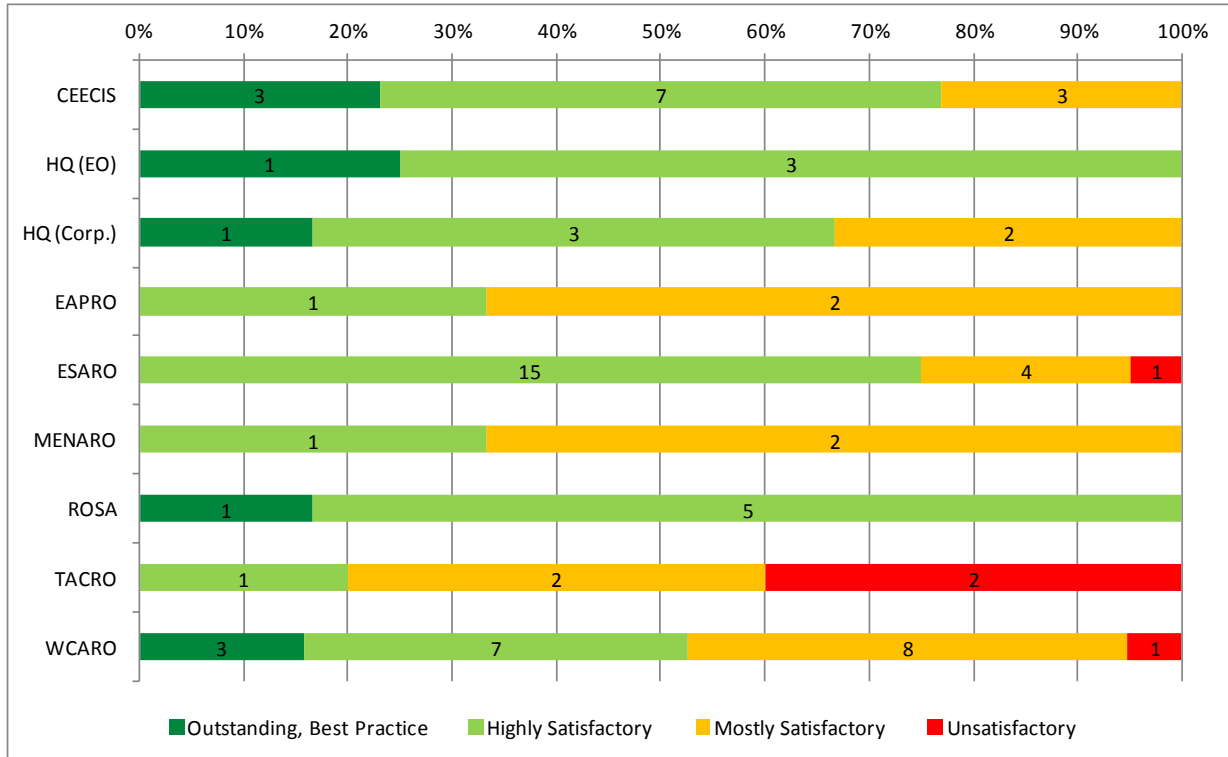
Section D Ratings



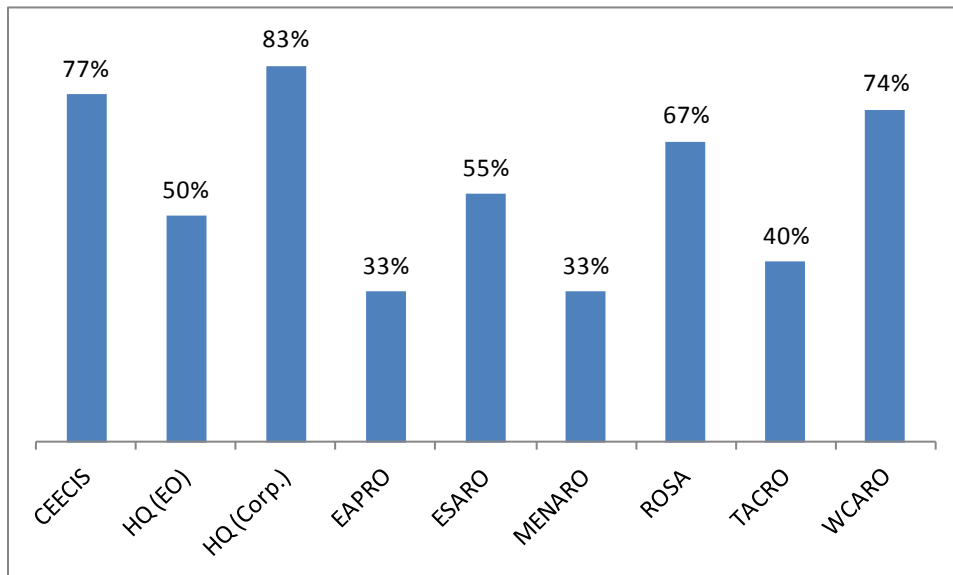
Section E Ratings



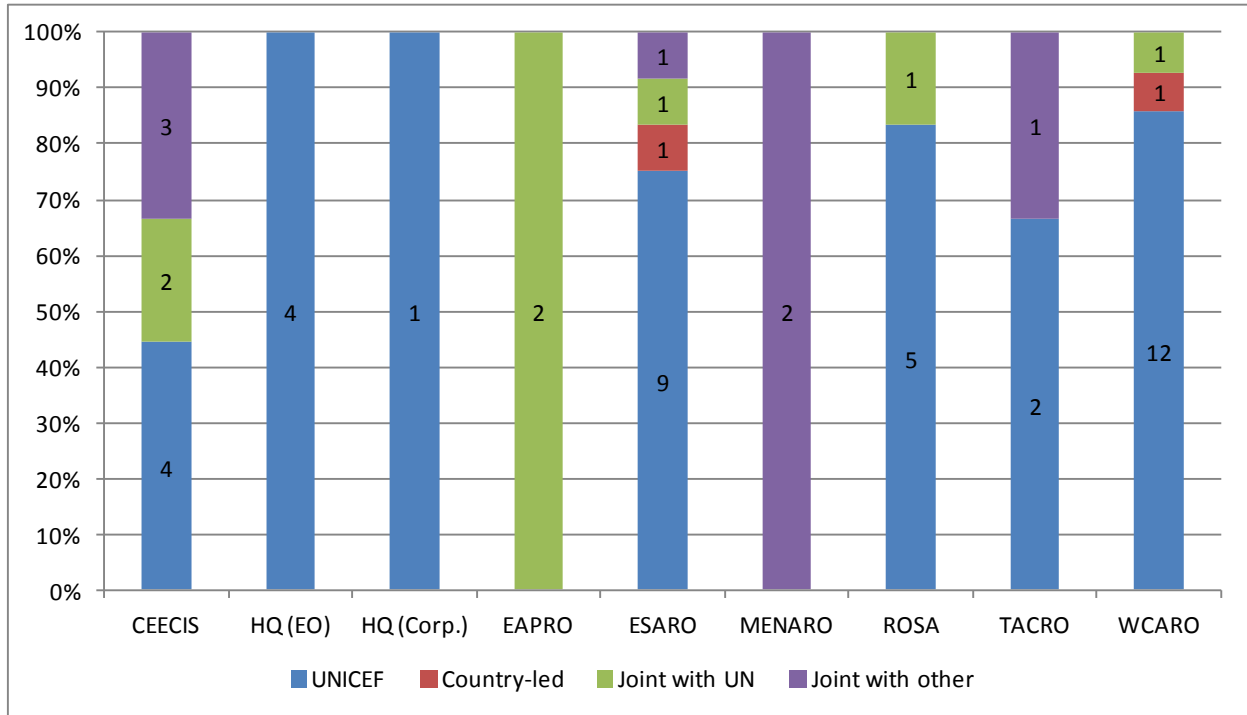
Section F Ratings



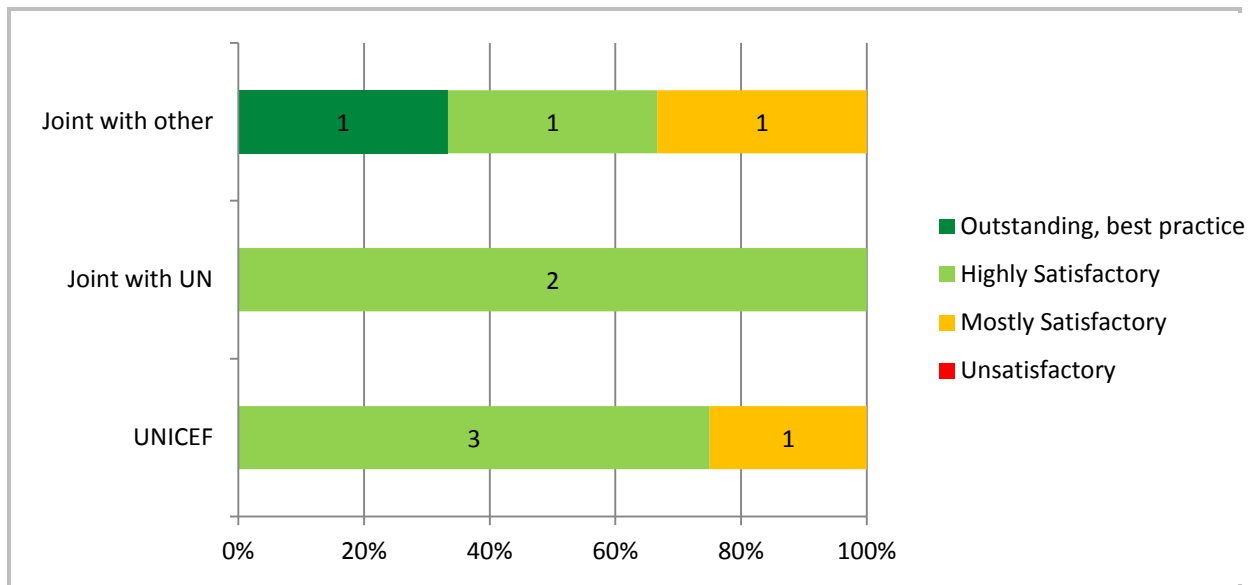
Inclusion of ToRs by Region for 2012



Appendix VII UNICEF, Country-led and Jointly Managed Evaluations by Region

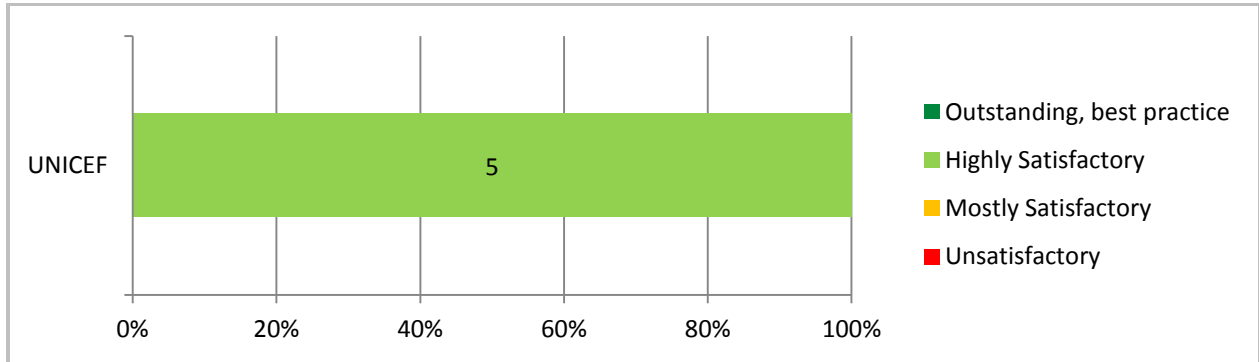


CEE/CIS



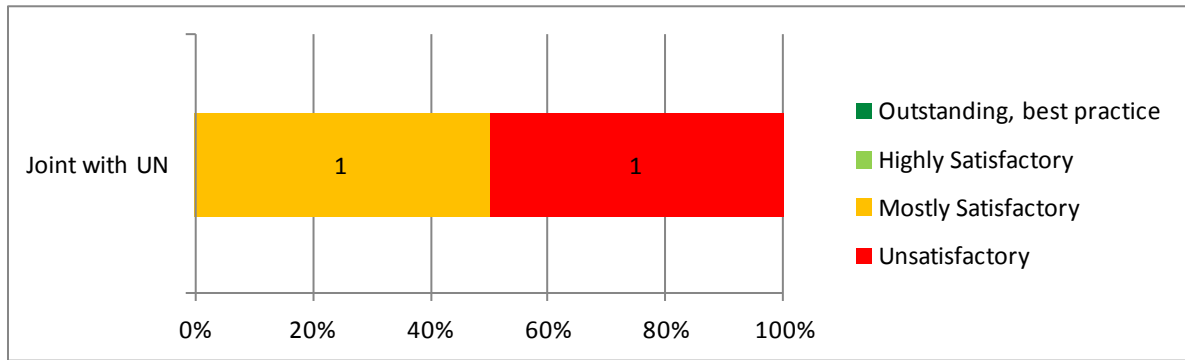
* No country-led evaluations in the region

HQ (EO and Corporate)



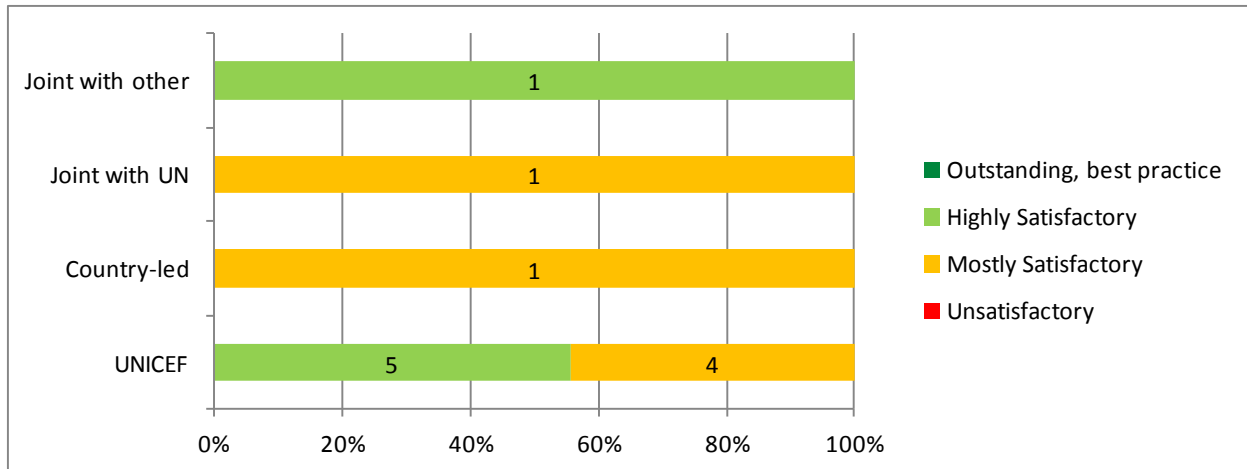
* For HQ (EO-led and corporate), there are no country-led nor joint evaluations

EAPRO

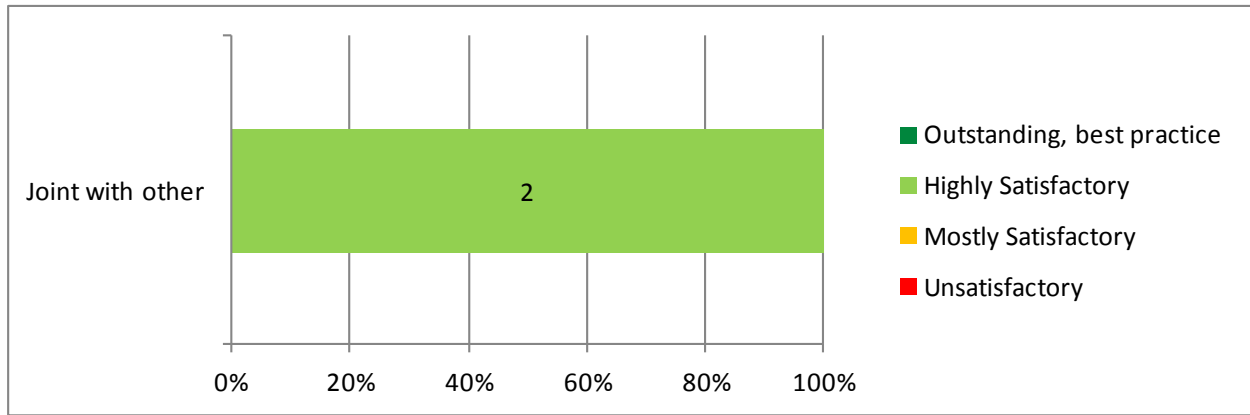


* For EAPRO, there are no UNICEF managed, country-led or joint with other evaluations.

ESARO

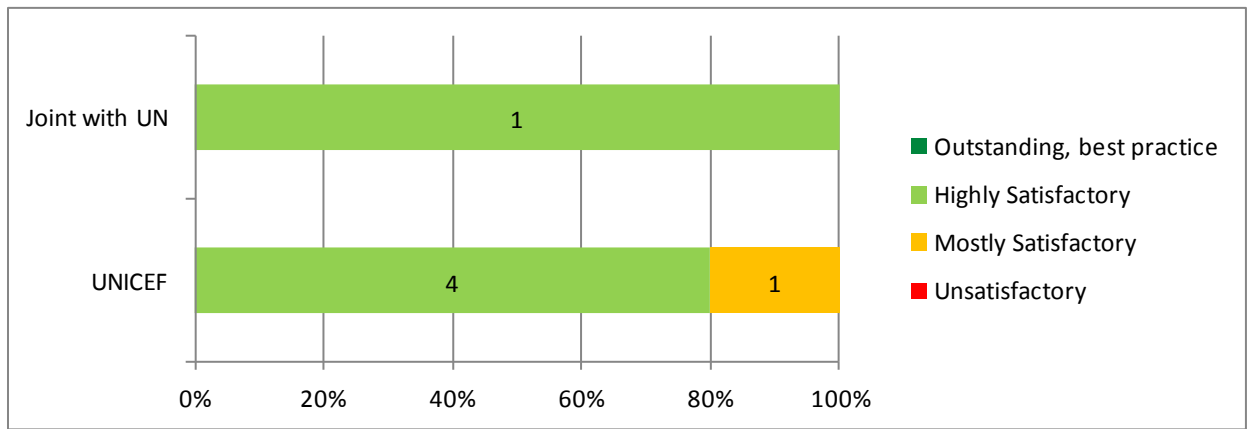


MENARO



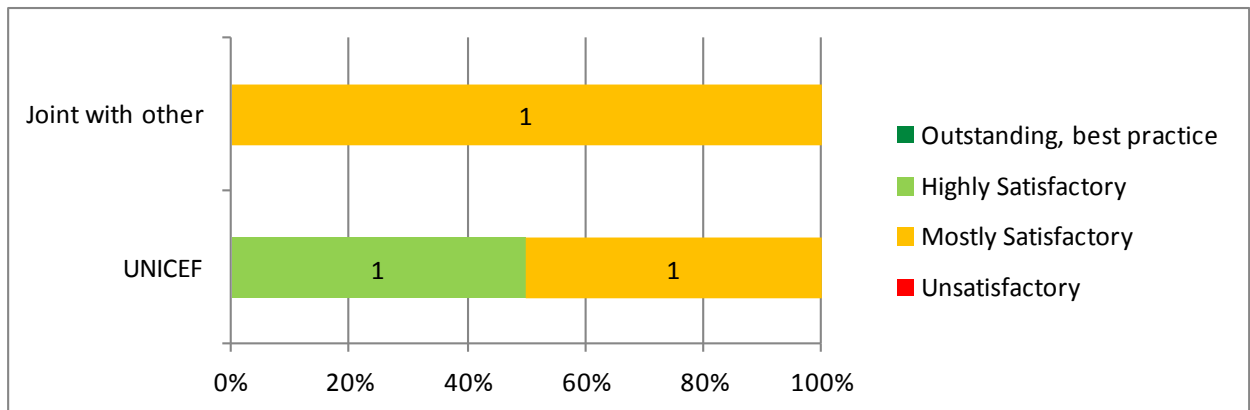
* For MENARO, there are no UNICEF-led, country-led nor jointly managed with UN evaluations

ROSA



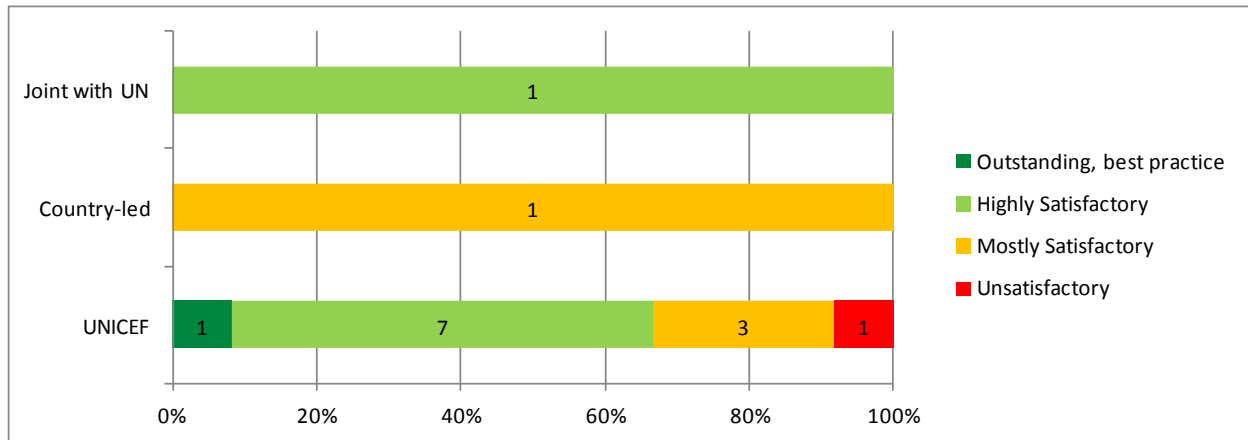
* For ROSA , there are no country-led nor Jointly managed with other evaluations

TACRO



* For TACRO, there are no country-led nor Jointly managed with UN evaluations

WCARO



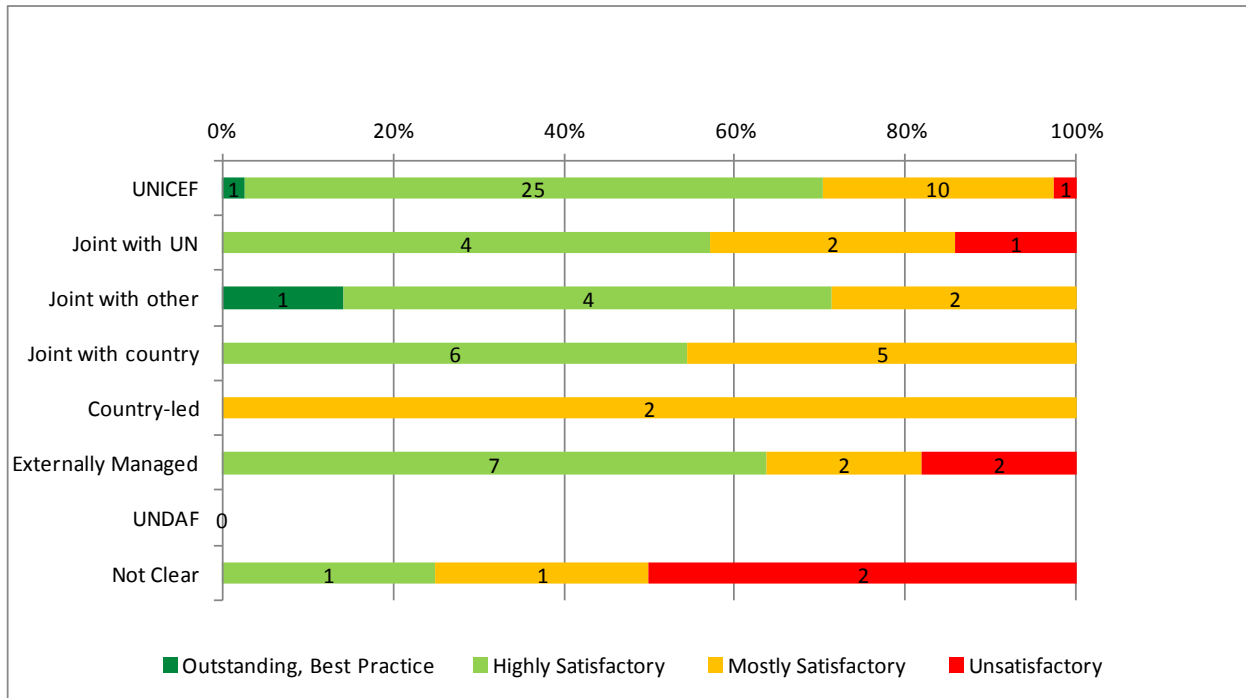
* For WCARO , there are no Jointly managed with others evaluations

Appendix VIII Overall Ratings Graphs by Report Typology

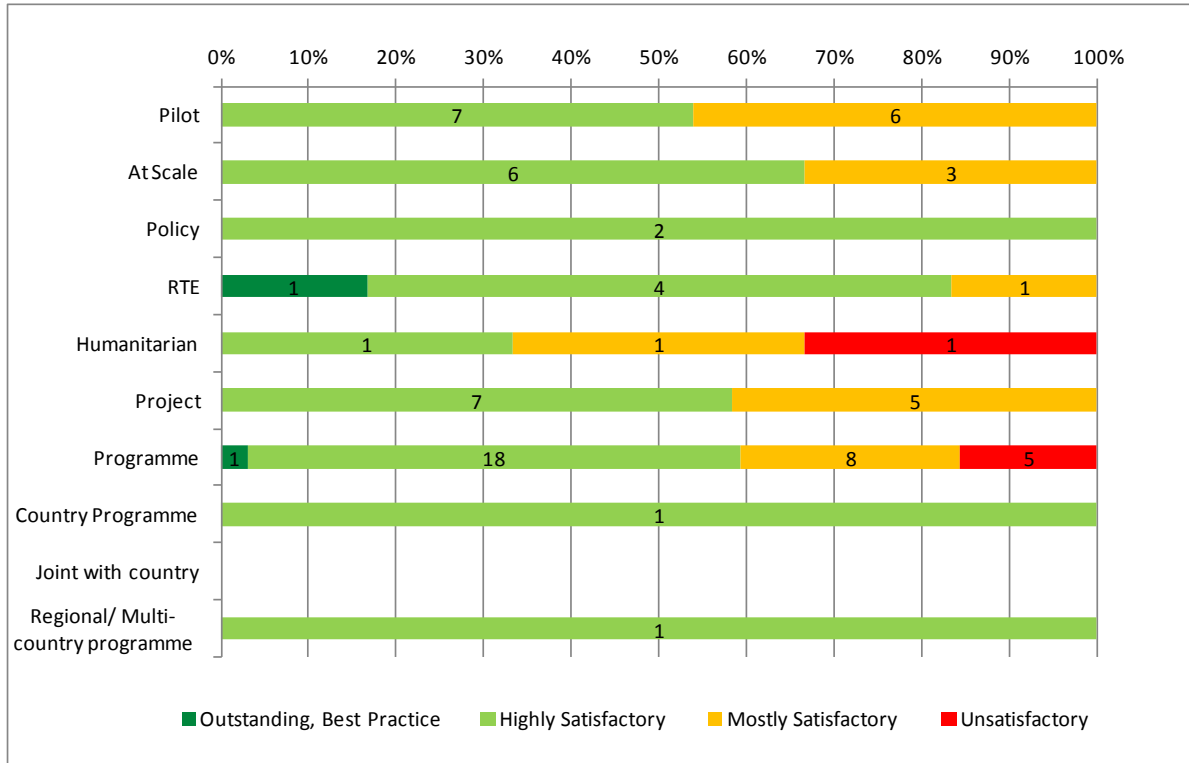
Geography



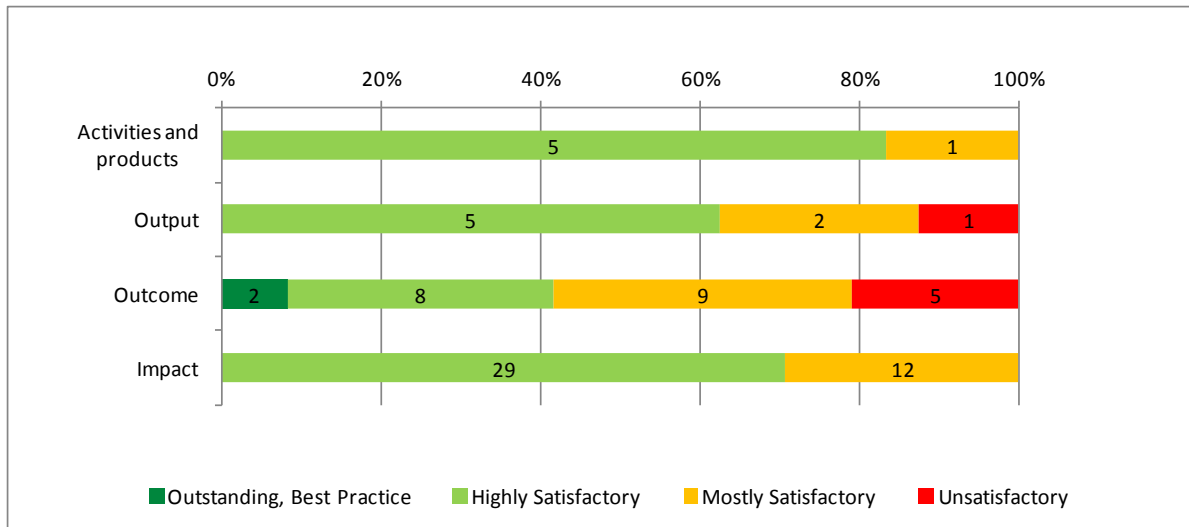
Management of the evaluation



Purpose of the Evaluation



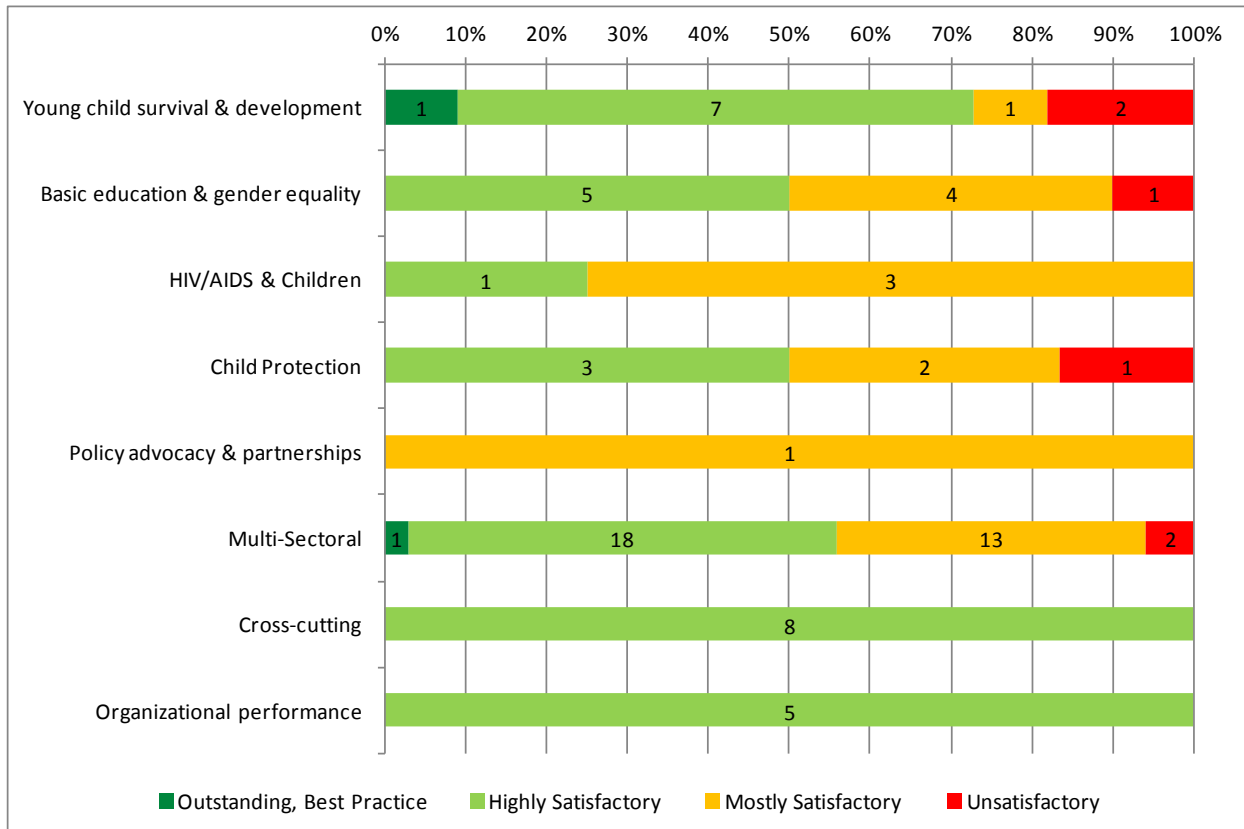
Results Level



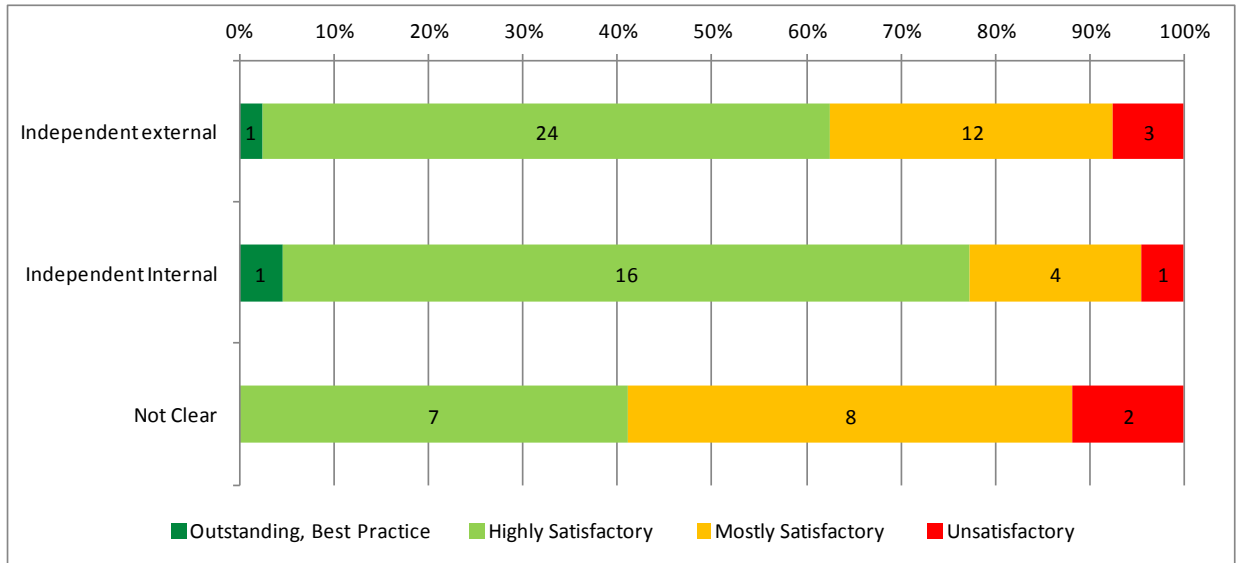
Stage



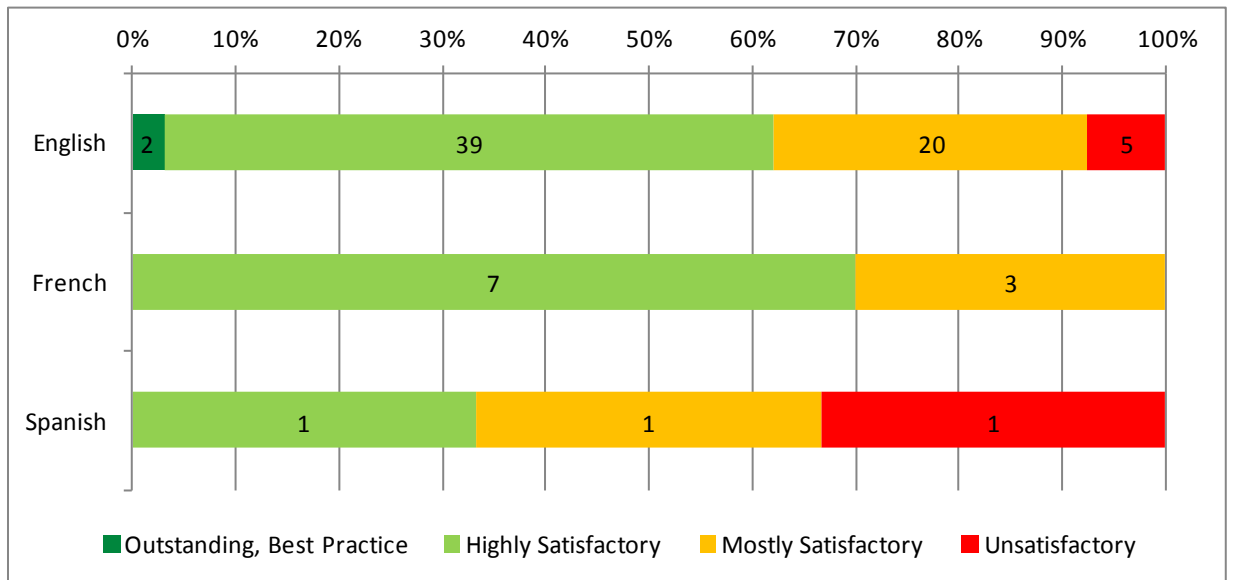
MTSP Correspondence



Level of independence

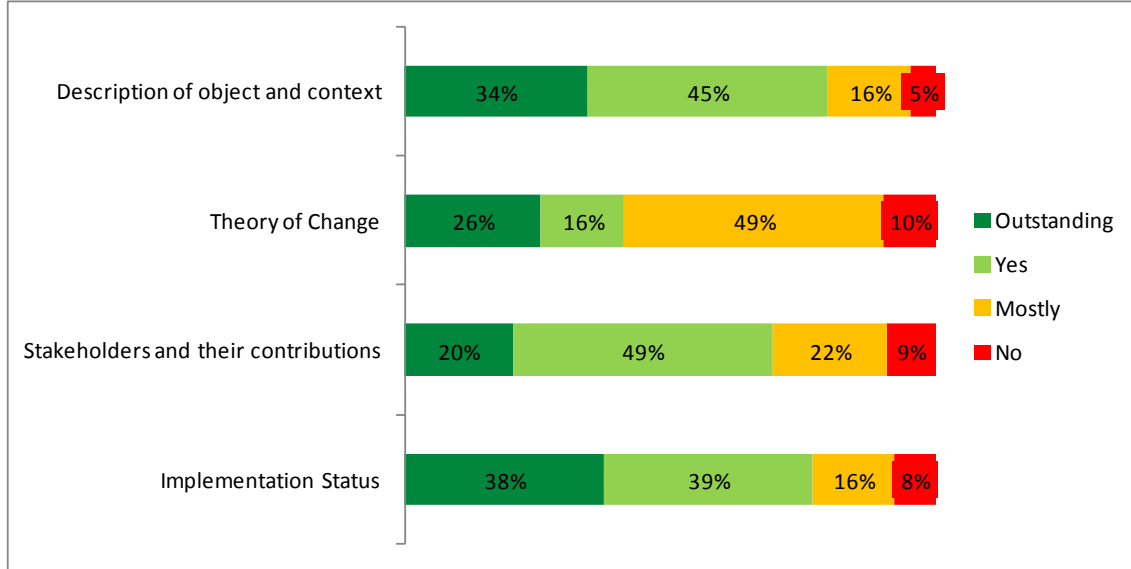


Language

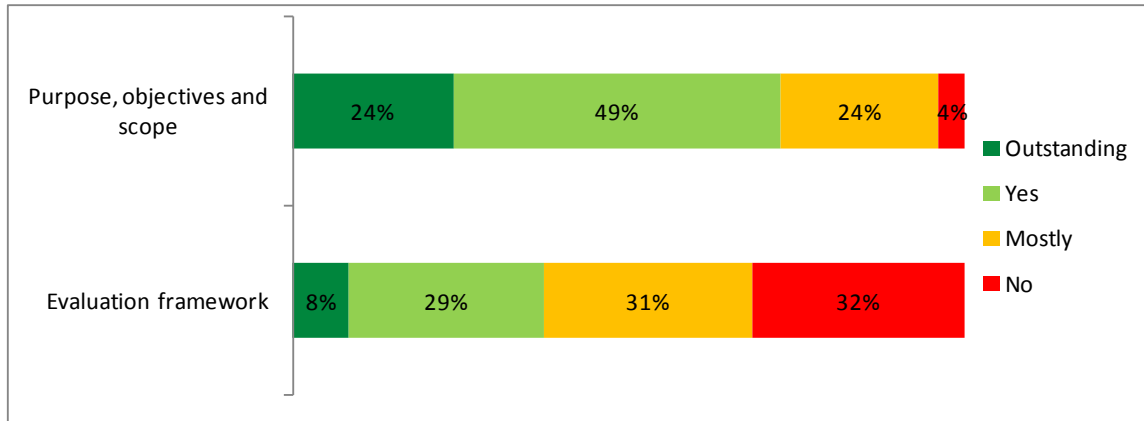


Appendix IX Sub-section ratings

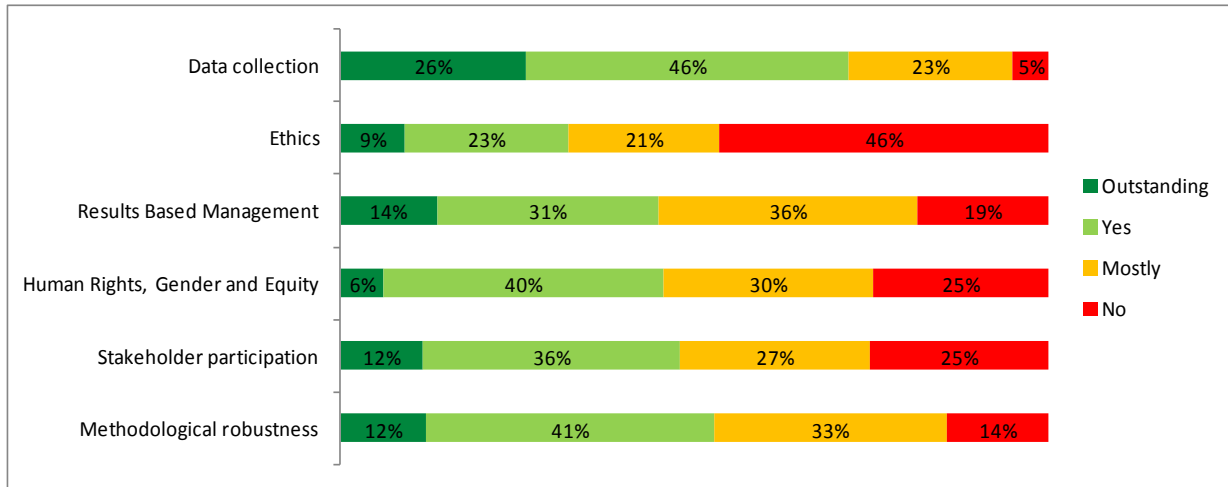
Section A



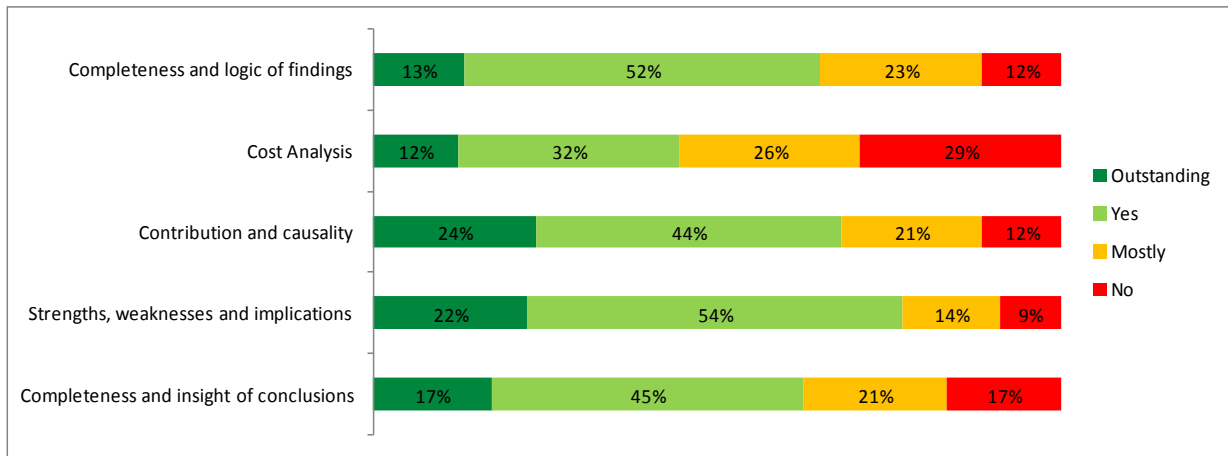
Section B



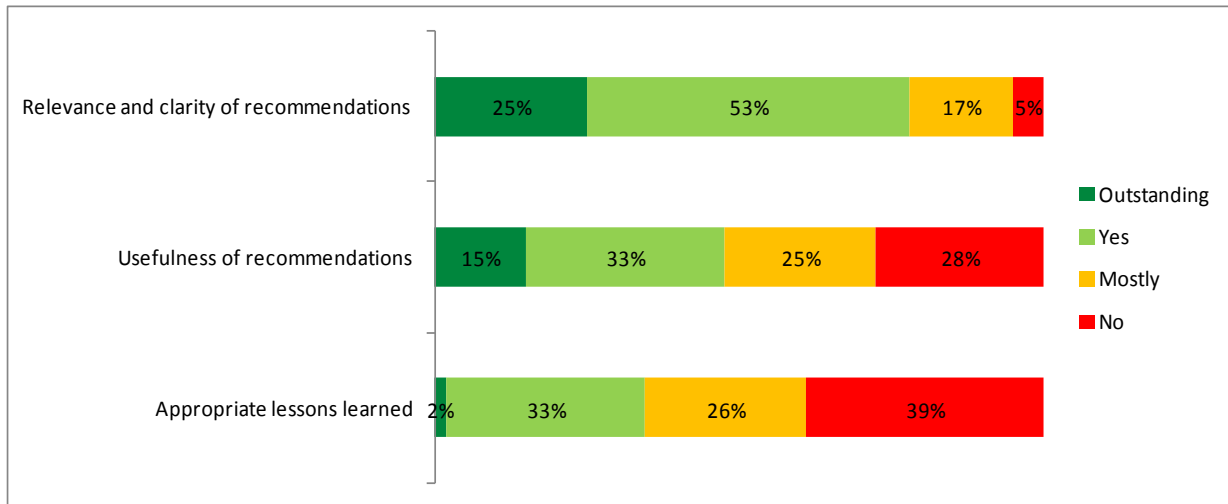
Section C



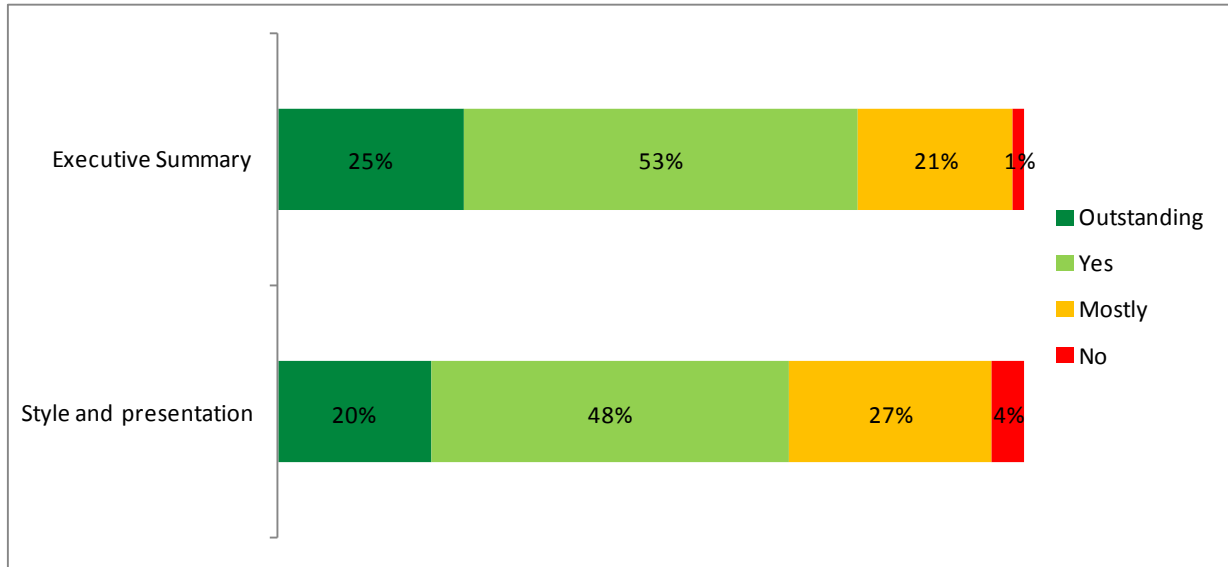
Section D



Section E



Section F



Appendix X Analysis Table for 2012 Reports

Rating Category:	Regional Trends									Geography					Management						Purpose							Result				Level of independence				Stage			MSTP Correspondence												
	CEECIS	HQ (EO)	HQ (Corp.)	EAPRO	ESARO	MENARO	ROSA	TACRO	WCARO	Sub-national	National	Multi-country	Regional	Multi-region/global	UNICEF	Joint with UN	Joint with other	Joint with country	Country-led	Externally Managed	UNDAF	Not Clear	Pilot	At Scale	Policy	RTE	Humanitarian	Project	Programme	Country Programme	Joint with country	Regional/ Multi-	Activities and	Output	Outcome	Impact	Self-evaluation	Independent internal	Independent external	Not clear	Formative	Summative	Formative and	Young child survival &	Basic education &	HIV/AIDS & Children	Child Protection	Policy advocacy &	Multi-Sectoral	Cross-cutting	Organizational
OVERALL																																																			
Outstanding, Best Practice	1	0	0	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	2	0	0	1	1	0	1	1	0	1	0	0	0	1	0	0
Highly Satisfactory	9	4	5	1	10	2	5	1	10	17	23	0	2	5	25	4	4	6	0	7	0	1	7	6	2	4	1	7	18	1	0	1	5	5	8	29	0	16	24	7	12	17	18	7	5	1	3	0	18	8	5
Mostly Satisfactory	2	0	1	1	9	1	1	2	7	6	16	2	0	0	10	2	2	5	2	2	0	1	6	3	0	1	1	5	8	0	0	0	1	2	9	12	0	4	12	8	5	8	11	1	4	3	2	1	13	0	0
Unsatisfactory	1	0	0	1	1	0	0	2	1	3	3	0	0	0	1	1	0	0	0	2	0	2	0	0	0	0	1	0	5	0	0	0	0	1	5	0	0	1	3	2	1	5	0	2	1	0	1	0	0		
SECTION A																																																			
Outstanding, Best Practice	5	2	1	0	2	0	0	1	3	3	5	1	2	3	10	0	2	1	0	1	0	0	3	1	0	1	1	4	4	0	0	0	1	1	3	9	0	7	6	1	5	4	5	3	1	0	0	6	2	2	
Highly Satisfactory	6	1	3	1	13	1	6	2	8	12	26	1	1	1	19	5	4	7	0	6	0	0	5	5	2	3	0	5	20	1	0	0	3	3	9	26	0	10	22	9	8	16	17	6	4	2	4	0	17	6	2
Mostly Satisfactory	2	1	2	2	4	2	0	2	6	8	11	1	0	1	6	2	1	3	2	3	0	4	5	3	0	2	0	3	7	0	0	1	2	4	9	6	0	4	10	7	6	9	6	2	5	2	2	1	8	0	1
Unsatisfactory	0	0	0	0	1	0	0	0	2	3	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	2	0	1	0	0	0	0	0	0	3	0	0	1	2	0	0	2	1	0	0	0	0	3	0	0	
SECTION B																																																			
Outstanding, Best Practice	2	1	0	0	1	0	0	0	0	1	1	1	0	1	2	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	1	0	2	1	0	1	2	1	2	1	1	1	1	0	0	0	2	0	1	
Highly Satisfactory	7	2	2	1	9	1	5	2	10	12	21	1	2	3	22	3	3	5	1	5	0	0	4	5	1	2	1	5	20	1	0	0	2	3	8	26	0	14	17	8	11	11	17	5	6	2	2	0	14	7	3
Mostly Satisfactory	4	1	4	1	8	2	1	2	6	10	16	1	1	1	13	3	3	3	0	4	0	3	9	1	0	4	2	5	7	0	0	1	2	5	10	12	0	6	18	5	4	15	10	4	3	2	3	1	14	1	1
Unsatisfactory	0	0	0	1	2	0	0	1	3	3	4	0	0	0	0	0	0	2	1	2	0	1	0	2	0	0	0	1	4	0	0	0	1	0	4	2	0	1	3	3	2	4	1	1	1	0	1	0	4	0	0

Rating Category:	Regional Trends								Geography					Management						Purpose							Result				Level of independence				Stage			MSTP Correspondence													
	CEECIS	HQ (EO)	HQ (Corp.)	EAPRO	ESARO	MENARO	ROSA	TACRO	WCARO	Sub-national	National	Multi-country	Regional	Multi-region/global	UNICEF	Joint with UN	Joint with other	Joint with country	Country-led	Externally Managed	UNDAF	Not Clear	Pilot	At Scale	Policy	RTE	Humanitarian	Project	Programme	Country Programme	Joint with country	Regional/ Multi-	Activities and	Output	Outcome	Impact	Self-evaluation	Independent internal	Independent external	Not clear	Formative	Summative	Formative and	Young child survival &	Basic education &	HIV/AIDS & Children	Child Protection	Policy advocacy &	Multi-Sectoral	Cross-cutting	Organizational
SECTION C																																																			
Outstanding, Best Practice	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Highly Satisfactory	7	4	2	2	9	1	3	1	8	13	15	2	2	5	20	2	4	6	0	4	0	1	4	3	2	2	2	5	17	1	0	1	4	3	8	22	0	13	20	4	13	11	13	6	5	2	3	0	15	1	5
Mostly Satisfactory	4	0	3	0	10	2	3	3	11	11	24	0	1	0	17	3	3	5	2	5	0	1	9	5	0	3	1	7	11	0	0	0	2	3	13	18	0	9	16	11	5	16	15	4	4	2	2	1	17	6	0
Unsatisfactory	2	0	1	1	1	0	0	1	0	2	3	1	0	0	0	2	0	0	0	2	0	2	0	1	0	1	0	0	4	0	0	0	0	2	3	1	0	0	4	2	1	4	1	1	1	0	1	0	2	1	0
SECTION D																																																			
Outstanding, Best Practice	0	1	1	0	0	0	0	1	1	1	0	0	1	2	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	2	0	0	0	3	0	1	1	1	1	1	1	0	0	1	0	0	
Highly Satisfactory	10	1	3	1	11	2	5	1	9	14	24	1	2	2	23	3	5	7	0	5	0	0	5	5	1	3	1	7	19	1	0	1	1	4	8	30	0	16	18	9	10	15	18	7	4	2	3	0	17	8	2
Mostly Satisfactory	2	2	2	1	8	1	1	3	9	9	15	2	1	2	12	3	2	4	2	4	0	2	8	3	1	2	2	5	8	0	0	0	5	2	11	11	0	6	17	6	7	12	10	2	4	2	2	1	15	0	3
Unsatisfactory	1	0	0	1	1	0	0	1	0	2	2	0	0	0	0	1	0	0	0	1	0	2	0	0	0	0	0	4	0	0	0	0	1	3	0	0	0	2	2	1	3	0	1	1	0	1	0	1	0	0	
SECTION E																																																			
Outstanding, Best Practice	0	0	0	0	0	0	0	3	1	1	0	1	0	2	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	2	1	0	1	2	0	3	0	0	1	1	0	0	0	1	0	0	
Highly Satisfactory	6	2	4	1	10	0	4	0	9	12	17	2	2	3	21	1	2	6	1	4	0	1	7	5	2	3	1	5	12	1	0	0	4	4	8	20	0	14	14	8	9	16	11	7	3	1	0	0	16	5	4
Mostly Satisfactory	7	2	2	2	8	3	2	4	6	12	21	1	0	2	14	6	5	2	1	6	0	2	6	2	0	2	2	6	17	0	0	1	1	4	12	19	0	6	23	7	6	12	18	3	6	3	5	1	14	3	1
Unsatisfactory	0	0	0	0	2	0	0	1	1	1	3	0	0	0	0	0	0	2	0	1	0	1	0	1	0	0	0	3	0	0	0	1	0	2	1	0	1	1	2	1	3	0	0	0	0	1	0	3	0	0	

Rating Category:	Regional Trends								Geography					Management						Purpose							Result				Level of independence				Stage			MSTP Correspondence													
	CEECIS	HQ (EO)	HQ (Corp.)	EAPRO	ESARO	MENARO	ROSA	TACRO	WCARO	Sub-national	National	Multi-country	Regional	Multi-region/global	UNICEF	Joint with UN	Joint with other	Joint with country	Country-led	Externally Managed	UNDAF	Not Clear	Pilot	At Scale	Policy	RTE	Humanitarian	Project	Programme	Country Programme	Joint with country	Regional/ Multi-	Activities and	Output	Outcome	Impact	Self-evaluation	Independent internal	Independent external	Not clear	Formative	Summative	Formative and	Young child survival &	Basic education &	HIV/AIDS & Children	Child Protection	Policy advocacy &	Multi-Sectoral	Cross-cutting	Organizational
SECTION F																																																			
Outstanding, Best Practice	3	1	1	0	0	0	1	0	3	2	4	1	1	1	5	0	2	1	0	1	0	0	1	2	0	2	0	1	3	0	0	0	0	1	3	5	0	5	3	1	4	3	2	2	3	0	0	0	3	1	0
Highly Satisfactory	7	3	3	1	15	1	5	1	7	15	23	0	1	4	23	5	2	8	0	4	0	1	10	5	2	2	1	6	15	1	0	1	4	4	8	27	0	13	24	6	11	17	15	7	2	2	4	0	18	6	4
Mostly Satisfactory	3	0	2	2	4	2	0	2	8	8	12	2	1	0	8	2	3	2	2	4	0	2	1	2	0	2	2	5	11	0	0	0	2	3	9	9	0	4	11	8	4	8	11	1	5	2	1	1	11	1	1
Unsatisfactory	0	0	0	0	1	0	0	2	1	1	3	0	0	0	1	0	0	0	0	2	0	1	1	0	0	0	0	0	3	0	0	0	0	0	4	0	0	0	2	2	0	3	1	1	0	0	1	0	2	0	0

Appendix XI Clarification of criteria for completing review

Region	The Americas and Caribbean Regional Office	Central & Eastern Europe, Commonwealth of Independent States RO	East Asia and the Pacific Regional Office	Eastern and Southern Africa Regional Office	Middle East and North Africa Regional Office	South Asia Regional Office	West and Central Africa Regional Office	Corporate (HQ)	Other		
Date of review											
Type of Report	Evaluation	Needs assessment	Appraisal	Evaluability	Review, including mid-term review	Inspection	Investigation	Research & study	Audit	Survey	Internal Management consulting
Geographical Coverage of the programme being evaluated and generalizability of evaluation findings	1.1 Sub-national: The programme and evaluation covers selected sub-national units (districts, provinces, states, etc.) within a country, where results cannot be generalized to the whole country	1.2 National: The programme covers the whole country, and the evaluation draws a sample in every district, or uses a sampling frame that is representative of the whole country.	1.3 Multi-country: Where one programme is implemented in several countries, or different programmes of a similar theme are implemented in several countries, the evaluation would cover two or more countries within one region. The results of the evaluation would not be generalizable to other countries in the region.	1.4 Regional: Where one programme is implemented in several countries, or different programmes of a similar theme are implemented in several countries, the evaluation covers multiple countries within the region and the sampling is adequate to make the results generalizable to the region.	1.5 Multi-region/Global: The programme is implemented in two or more regions, or deliberately targets all regions. The evaluation would typically sample several countries across multiple regions, with the results intended to be generalizable in two or more regions.						
Management Managerial control and oversight of evaluation decisions (i.e., ToRs, selection of consultants, budgets, quality assurance and approval of evaluation findings). In all instances, it is assumed that the management approaches include relevant national actors (e.g., government, universities, NGOs, CBOs)	2.1 UNICEF managed: Working with national partners of different categories UNICEF is responsible for all aspects of the evaluation.	2.2 Joint managed, with one or more UN agencies: UNICEF is the co-manager with one or more UN agencies	2.3 Joint managed, with organisations outside the UN system: UNICEF is the co-manager with one or more organizations outside the UN system	2.4. Jointly Managed with Country: Evaluations jointly managed by the Country (Government and/or CSO) and the UNICEF CO	2.5. Country-led Evaluation: Evaluations managed by the Country (Government and/or CSO)	2.6 UNDAF Evaluation: They are a joint UN review, conducted with national partners, about the overall results expected from the UN cooperation in the country.	2.7 Externally managed: An external organization manages the evaluation, where UNICEF is one of the organizations being assessed (UN and non-UN)	2.8 Not clear from Report			

Region	The Americas and Caribbean Regional Office	Central & Eastern Europe, Commonwealth of Independent States RO	East Asia and the Pacific Regional Office	Eastern and Southern Africa Regional Office	Middle East and North Africa Regional Office	South Asia Regional Office	West and Central Africa Regional Office	Corporate (HQ)	Other		
Purpose <i>Speaks to the overarching goal for conducting the evaluation; its raison d'être</i>	3.1 Pilot: Where a new solution, approach, or programme is being tested at a national or sub-national level, the evaluation examines the efficacy of such an intervention with the intention to determine suitability for scaling-up.	3.2 At scale: The evaluation examines the efficacy of a programme that is being implemented at or near its maximum intended extent, with the intention of providing feedback on efficiency and the overall effectiveness of the programme to scale up focus for lessons learned.	3.3 Policy: An evaluation whose main purpose is to examine the results of a policy that is delinked from field-based programming operations.	3.4 Real-time-evaluation: In the context of an emergency, an evaluation of the efficacy of the response, which collates lessons that can be applied back to an on-going response	3.5 Humanitarian: Humanitarian evaluation assesses organizational performance in emergency settings (including both natural disasters & conflicts) at various phases of these crises, from preparedness and risk reduction to response, recovery & the transition to development	3.6 Project: An evaluation which is step-by-step process of collecting, recording and organisation information about the project results including immediate results, short-term outputs and long-term project outcomes	3.7 Programme: An evaluation of a sectorial programme to determine its overall effectiveness and efficiency in relation to the stated goals and objectives	3.8 Country Programme Evaluation (CPE): An evaluation that assess the relevance, effectiveness, efficiency, sustainability of the entire UNICEF Country Programme	3.9 Regional/ Multi-country programme evaluation: An evaluation that assesses several programmes from a regional or multi-country perspective		
Result <i>Level of changes sought, as defined in results based management: refer to substantial use of highest level reached</i>	4.1 Output: Causal effects deriving directly from programme activities, and assumed to be completely under programme control	4.2 Outcome: Effects from one or more programmes being implemented by multiple actors (UNICEF and others), where the cumulative effect of outputs elicits results beyond the control of any one agency or programme	4.3 Impact: Final results of a programme or policy on the intended beneficiaries and, where possible, on comparison groups. Reflects the cumulative effect of donor supported programmes of cooperation and national policy initiatives.	4.4 Activities and products: Describes things that have been done rather than their effects (workshops given, publications produced, meetings attended or organized)							
MTSP Correspondence <i>Alignment with MTSP focus area priorities: (1) Young child survival and development; (2) Basic education and gender equality; (3) HIV/AIDS and children; (4) Child protection from violence, exploitation and abuse; and (5) Policy advocacy and partnerships for children's rights</i>	5.1 Sectoral: addresses issues within only one of the five MTSP focus areas (1. Young child survival & development)	5.1 Sectoral: addresses issues within only one of the five MTSP focus areas (2. Basic education & gender equality)	5.1 Sectoral: addresses issues within only one of the five MTSP focus areas (3. HIV/AIDS & children)	5.1 Sectoral: addresses issues within only one of the five MTSP focus areas (4. Child Protection)	5.1 Sectoral: addresses issues within only one of the five MTSP focus areas (5. Policy advocacy & partnerships)	5.2 Multi-sectoral: Addresses issues in two or more MTSP focus areas	5.3 Cross-cutting: Addresses issues that are named as cross-cutting strategies of the MTSP or otherwise known to operate within all MTSP areas. Includes but is not limited to the human rights-based approach to programming, gender equity, knowledge management, evaluation, & communication	5.4 Organizational performance: Evaluation of institutional effectiveness in operational areas, including supply, information technology, human resources, and finance. Also includes the more general issue of overall quality of UNICEF planning and implementation			

Region	The Americas and Caribbean Regional Office	Central & Eastern Europe, Commonwealth of Independent States RO	East Asia and the Pacific Regional Office	Eastern and Southern Africa Regional Office	Middle East and North Africa Regional Office	South Asia Regional Office	West and Central Africa Regional Office	Corporate (HQ)	Other		
							for development.				
Level of Independence <i>Implementation and control of the evaluation activities</i>	6.1 Self-evaluation: A significant component of evaluation management activities & decision-making about the evaluation are implemented by individuals associated with the target programme/intervention (eg. programmes officer/specialists)	6.2 Independent internal: The evaluation is implemented by consultants but managed in-house by UNICEF professionals. The overall responsibility for the evaluation lies within the division whose work is being evaluated.	6.3 Independent external: The evaluation is implemented by external consultants and/or UNICEF Evaluation Office professionals. The overall responsibility for the evaluation lies outside the division whose work is being evaluated.	6.4 Not clear from Report							
Stage	7.1 Formative: An evaluation with the purpose and aim of improving the programme. Formative evaluations strengthen or improve the object being evaluated by examining the delivery of the programme	7.2 Summative: An evaluation that examines the effects or outcomes of the object being evaluated and summarize it by describing what happened subsequent to delivery of the programme	7.3 Summative and formative: An evaluation that combines the elements of a formative and a summative evaluation.								
ToRs present	Yes	No									
Question Criteria	Outstanding	Yes	Mostly	No	N/A						
Section Rating Criteria	Outstanding, best practice	Highly satisfactory	Mostly Satisfactory	Unsatisfactory							